UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO - USP
DEPARTAMENTO DE GENÉTICA

BRUNA RODRIGUES MUYS

“Caracterização da estrutura e regulação dos genes

***MGC16121 e CR596471*”**

RIBEIRÃO PRETO-SP

2013

BRUNA RODRIGUES MUYS

**“Caracterização da estrutura e regulação dos genes
MGC16121 e *CR596471*”**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Genética da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, para a obtenção do título de Mestre em Ciências Biológicas-Área de Genética.

Orientador: Prof. Dr. Wilson Araújo da Silva Júnior

RIBEIRÃO PRETO-SP

2013

Muys, Bruna Rodrigues

Caracterização da estrutura e regulação dos genes *MGC16121* e *CR596471*.

Ribeirão Preto, 2013.

122p.; 30cm

Dissertação de Mestrado apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo. Departamento de Genética. Área de concentração: Genética.

Orientador: Wilson Araújo da Silva Júnior.

1- câncer; 2- expressão gênica; 3- regulação por metilação; 4- RNAs não codificadores; 5- biomarcadores; 6- genes *MGC16121* e *CR596471*.

FOLHA DE APROVAÇÃO

Bruna Rodrigues Muys

Caracterização da estrutura e regulação dos genes *MGC16121* e *CR596471*

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Genética da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, para a obtenção do título de Mestre em Ciências Biológicas.

Área de Concentração: Genética

Aprovado em: ____/____/____.

Banca Examinadora

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

Prof. Dr. _____

Instituição: _____ Assinatura: _____

"A vida sem ciência é
uma espécie de morte."

Sócrates

Dedico aos meus pais,
Maura e Boudewijn,
que sempre se
dedicaram e me
incentivaram ao longo
da minha vida.

AGRADECIMENTOS

Ao Professor Dr. Wilson Araújo da Silva Júnior, o qual me deu confiança, oportunidade de trabalho e aprendizado;

Aos amigos do laboratório (Rafaela, Dalila, Júlio, Nathália, Willys, Luíza, Karina, Aline, Anemari, Manuela, Cristiane, Greice, Cleidson, Frederico, Carolina, Ana Júlia, Adriana, Meire, Marcelo e Daniéis) pela amizade, orientação, atenção e compreensão, que tornaram a realização cotidiana deste projeto agradável e inesquecível;

Aos familiares que sempre me apoiaram durante a realização deste trabalho;

Aos docentes e funcionários do Hemocentro de Ribeirão Preto e Departamento de Genética da Faculdade de Medicina de Ribeirão Preto;

Ao CNPq e a FAPESP, pelo apoio financeiro concedido para a realização deste trabalho;

A todos os outros amigos e colegas que direta ou indiretamente contribuíram de algum modo para minha formação e concretização deste trabalho.

RESUMO

MUYS, BR. (2013). Caracterização da estrutura e regulação dos genes *MGC16121* e *CR596471*. Dissertação de Mestrado, Faculdade de Medicina, Universidade de São Paulo, Ribeirão Preto.

Os genes *MGC16121* e *CR596471* localizam-se no cromossomo X (Xq26) entre os loci *HPRT1* e *PLAC1*, uma região rica em genes associados com a reprodução humana. A importância de tais genes reside na possibilidade de estarem envolvidos no desenvolvimento placentário e fetal e de serem expressos em poucos tecidos normais. Camundongos portadores de deleções próximas do gene ortólogo *HPRT1* de humanos apresentam cerca de um terço do tamanho dos camundongos selvagens ou em alguns casos são natimortos. No entanto, este fenótipo não é observado quando o gene está mutado. Assim, pode-se supor que o fenótipo anormal das cobaias não é resultado da deficiência do *HPRT1*, mas sim de genes e/ou microRNAs (miRNAs) próximos a ele. Estes resultados abrem perspectivas em relação ao estudo dos genes *MGC16121*, *CR596471* e miRNAs das vizinhanças. O objetivo deste trabalho foi caracterizar a estrutura, a expressão e o mecanismo de regulação por metilação dos genes *MGC16121* e *CR596471*. Adicionalmente foram analisados quanto ao perfil de expressão e regulação por metilação os miRNAs das vizinhanças (miR-424, 503, 450a, 450b-5p e 542-3p). O gene *MGC16121* mostrou-se específico de placenta e também expresso em 50% das 18 linhagens tumorais analisadas. Já *CR596471* e os miRNAs das vizinhanças foram mais expressos em placenta do que qualquer outro tecido normal analisado, sendo o primeiro expresso também em 100% das linhagens tumorais avaliadas. Houve correlação positiva e significativa entre todos os genes e miRNAs em relação à expressão em tecidos normais, porém o mesmo não foi observado para linhagens tumorais. A respeito da regulação, os genes *CR596471* e *MGC16121* e os miRNAs miR-424, 503 e 450a foram regulados negativamente por metilação do DNA em pelo menos uma das três linhagens tratadas com o agente demetilante 5-aza-2-deoxicitidina. Apoiando este fato, os dinucleotídeos CpG das ilhas CpGs situadas próximas às regiões 5' dos genes *CR596471* e *MGC16121* foram pelo menos em parte desmetilados após o mesmo tratamento. Os dados relativos à estrutura primária dos genes indicam que os transcritos, apesar de serem lncRNAs apresentaram características de mRNAs. Para *MGC16121* foi determinado um transcrito composto de 3 éxons e, para *CR596471*, um transcrito composto de 3 éxons e outro composto de 2 éxons. Os transcritos aqui determinados são relativamente conservados quando comparados a sequências de RNA encontradas em outros mamíferos, principalmente em primatas. Adicionalmente, o transcrito de *MGC16121* possui subestruturas secundárias visivelmente semelhantes com aquelas dos transcritos homólogos encontrados em alguns primatas. De acordo com os resultados, o gene *MGC16121* pode ser considerado um possível bom marcador para diagnóstico, prognóstico e talvez para terapias contra cânceres. Todavia, mais experimentos devem ser realizados para verificar a função dos genes *MGC16121* e *CR596471*, além de avaliar mais robustamente a capacidade do gene *MGC16121* ser utilizado como ferramenta na medicina contra o câncer.

Palavras-chave: câncer; expressão gênica; RNAs não codificadores; biomarcadores, genes *MGC16121* e *CR596471*.

ABSTRACT

MUYS, BR. (2013). Structural and regulatory characterization of genes *MGC16121* and *CR596471*. Dissertação de Mestrado, Faculdade de Medicina, Universidade de São Paulo, Ribeirão Preto.

CR596471 and *MGC16121* genes lie on chromosome X (Xq26) between the *HPRT1* and *PLAC1* loci, a region rich in genes associated with human reproduction. The importance of such genes is the possibility that they might be involved in placental and fetal development, aware that they are expressed in few normal tissues. Deletions in mice around the orthologous gene of human *HPRT1* affect their development or lead to stillbirth. However, this phenotype is not observed when this gene is mutated. So we can assume that the abnormal phenotype of mice cannot be due to *HPRT1* deficiency, but to genes and/or microRNAs (miRNAs) nearby. These results support the idea of investigating the mechanisms involved in the regulation of the *MGC16121* and *CR596471* genes, and their neighbor miRNAs. This study aimed to characterize the structure, expression and regulation mechanism by methylation of genes *MGC16121* and *CR596471*. In addition, the expression profile and methylation regulation of the neighbor miRNAs (miR-424, 503, 450a, 450b-5p and 542-3p) were analyzed. *MGC16121* was demonstrated to be placenta specific and expressed in 50% of 18 tumor cell lines analyzed. *CR596471* and the neighbor miRNAs were more expressed in placenta than in any other normal tissue analyzed. The former was also expressed in all tumor cell lines evaluated. There was significant and positive correlation between all genes and miRNAs regarding normal tissue expression. However, the same was not observed for the tumor cell lines. With respect to regulation, the genes *CR596471* and *MGC16121*, and miRNAs miR-424, 503 and 450a were negatively regulated by DNA methylation at least in one of the three cell lines treated with the demethylating agent 5-aza-2-deoxycytidine. Supporting these results, the CpG dinucleotides from CpG islands located near the *CR596471* and *MGC16121* 5' regions were at least partially demethylated after the same treatment. The data concerning to gene's primary structures indicate that the transcripts, despite of being considered lncRNAs, presented mRNAs characteristics. It was determined one transcript for *MGC16121* gene which consisted of three exons, and for *CR596471* gene, two transcripts were found, one with three exons and other composed of two exons. The transcripts herein determined are relatively conserved when compared to RNAs sequences found in other mammals, mostly in primates. Besides, the *MGC16121* transcript presents similar secondary substructures to those found in homologous transcripts from other primate species. According to the results, *MGC16121* gene could be considered a possible good biomarker to diagnosis, prognosis and perhaps to therapies against cancers. Nevertheless, more experiments must be accomplished in order to verify the functions of *MGC16121* and *CR596471* genes, in addition to evaluate more robustly the competence of *MGC16121* gene to be used as a tool in medicine against cancer.

Keywords: cancer; gene expression; non-coding RNAs; biomarkers; *MGC16121* and *CR596471* genes.

ABREVIATURAS

‰: por cento

5-Aza-dC: 5-Aza-2'-deoxicitidina

Airn: Antisense of IGF2R Non-protein coding RNA

ANOVA: *Analysis of Variance*

Antgenos CT: Antgenos Cncer-Testculo

BLAST: *Basic Local Alignment Search Tool*

BLAT: *BLAST-like alignment tool*

bp: Pares de base

cDNA: DNA complementar

CIP: Fosfatase Alcalina Intestinal de Bezerro

CpG: dinucletdeo CG

Ct: *cycle threshold*

DGCR8: DiGeorge syndrome critical region 8

DMSO: Dimetilsulfxido

DNA: *DeoxyriboNucleic Acid*

DNMT: DNA-metiltransferase

EGF: Epidermal Growth Factor

ENCODE: *Encyclopedia of DNA Elements*

ESTs: Etiquetas de Sequncia Expressa

FGF-7: Fibroblast Growth factor 7

G9a ou EHMT2: Euchromatic Histone-Lysine N-Methyltransferase 2

GAPDH: Glyceraldehyde 3-phosphate dehydrogenase

GSP: *Gene Specific Primer*

HDAC: Histona Deacetilase

HMT: Histona Metiltransferase

HOTAIR: HOX transcript antisense RNA

HOXD: Homeobox D

HP1: *Heterochromatin Protein 1*

Hprt: Hypoxanthine Phosphoribosyltransferase

HPRT1: Hypoxanthine Phosphoribosyltransferase 1

Igfr2: Fc fragment of IgG, low affinity IIa, receptor (CD32)

kb: Kilobase

Kcnq1: Potassium Voltage-gated Channel, KQT-like subfamily, member 1

Kcnq1ot1: KCNQ1 opposite strand/antisense transcript 1

LB: Luria Bertani

lncRNAs: RNAs longos não codificadores

LSD1 ou *KDM1A: Lysine (K)-Specific Demethylase 1A*

MAGE-A: *Melanoma Antigen Family A*

MAGE-C: *Melanoma Antigen Family C*

MBP: *Methyl CpG Binding Protein*

miRISC: *RNA-induced silencing complex* (RISC) e miRNA maduro

miRNAs: microRNAs

ml: mililitro

mRNAs: RNAs mensageiros

MS-HRM: *Methylation-Sensitive High Resolution Melting*

NAT: *Natural Antisense Transcripts*

nm: nanômetro

NY-ESO-1 ou *CTAG1B: Cancer/Testis Antigen 1B*

°C: grau Célsius

ORF: *Open Reading Frame*

PCR: *Polimerase Chain Reaction*

piRNAs: RNAs que interagem com proteínas Piwi

PLAC1: Placenta-Specific 1

PRC: *Protein Regulator of Cytokinesis*

PTEN: Phosphatase and Tensin Homolog

PTEN1: Phosphatase and Tensin Homolog Pseudogene 1

qRT-PCR: *Quantitative Reverse Transcriptase PCR*

QUMA: *Quantification Tool for Methylation Analysis*

RACE: *Rapid Amplification of cDNA Ends*

RNA: *RiboNucleic Acid*

RNase III: Ribonuclease III

rpm: Rotações Por Minuto

rRNAs: RNAs ribossômicos

siRNAs: pequenos RNAs de interferência

snoRNAs: pequenos RNAs nucleolares

snRNAs: pequenos RNA nucleares

SOC: *Super Optimal broth with Catabolite repression*

SPRY4-IT1: SPRY4 intronic transcript 1

SRA: *SET and RING finger-Associated*

TAP: Pirofosfatase Ácida de Tabaco

TRBP: Trans-Activation Response RNA Binding Protein

tRNAs: RNAs transportadores

UTR: *Untranslated Region*

V: volt

XIST: *X-inactive specific transcript*

μg: micrograma

μl: microlitro

μM: micromolar

LISTA DE FIGURAS

Figura 1. Semelhanças entre células germinativas/trofoblastos e células tumorais.....	22
Figura 2. Localização dos genes <i>MGC16121</i> e <i>CR596471</i>	34
Figura 3. Representação esquemática das atividades laboratoriais que foram realizadas no presente projeto.....	37
Figura 4. Esquema das primeiras etapas da técnica RACE para obtenção da extremidade 5'.....	40
Figura 5. Esquema da etapa de obtenção da primeira fita de cDNA.....	41
Figura 6. Esquema da etapa de obtenção da extremidade 5'	41
Figura 7. Esquema da etapa de obtenção da extremidade 3'	41
Figura 8. Ilhas CpG nas regiões 5' dos genes <i>MGC16121</i> e <i>CR596471</i>	45
Figura 9. qRT-PCR utilizando amostras de painel de tecidos normais.....	49
Figura 10. qRT-PCR utilizando amostras de painel de tecido normal para os miRNAs localizados próximos aos genes <i>MGC16121</i> e <i>CR596471</i>	50
Figura 11. Correlação de Pearson (r) entre diferentes loci em relação às expressões obtidas das amostras do painel de tecido normal.....	51
Figura 12. qRT-PCR utilizando amostras de linhagens tumorais.....	52
Figura 13. qRT-PCR utilizando amostras de linhagens tumorais para os miRNAs localizados próximos aos genes <i>MGC16121</i> e <i>CR596471</i>	52
Figura 14. Correlação de Pearson (r) entre diferentes loci em relação às expressões obtidas das amostras de linhagens tumorais.....	53
Figura 15. Correlação de Pearson (r) entre combinações de miRNAs que foram positivas em relação às expressões obtidas das amostras de linhagens tumorais.....	54
Figura 16. Comparação entre os níveis de expressão em tecidos normais e linhagens tumorais dos genes <i>MGC16121</i> e <i>CR596471</i> , e miRNAs da vizinhança.....	55
Figura 17. Comparação dos níveis de expressão dos genes <i>MGC16121</i> e <i>CR596471</i> , e miRNAs da vizinhança entre amostras de placentas obtidas em gestações de 37, 38 ou 40 semanas.....	56
Figura 18. Comparação dos níveis de expressão de <i>MGC16121</i> entre os grupos controle DMSO (veículo) e tratados com 5 μ M de 5-aza-2-deoxicitidina (5-Aza-dC) em três linhagens.....	57

Figura 19. Comparação dos níveis de expressão do miR-424 entre os grupos controle DMSO (veículo) e tratados com 5 µM de 5-aza-2-deoxicitidina (5-Aza-dC) em três linhagens.....	58
Figura 20. Comparação dos níveis de expressão do miR-503 entre os grupos controle DMSO (veículo) e tratados com 5 µM de 5-aza-2-deoxicitidina (5-Aza-dC) em três linhagens.....	59
Figura 21. Comparação dos níveis de expressão de <i>CR546971</i> entre os grupos controle DMSO (veículo) e tratados com 5 µM de 5-aza-2-deoxicitidina (5-Aza-dC) em três linhagens.....	60
Figura 22. Comparação dos níveis de expressão do miR-450a entre os grupos controle DMSO (veículo) e tratados com 5 µM de 5-aza-2-deoxicitidina (5-Aza-dC) em três linhagens.....	61
Figura 23. Padrão de metilação representativo de amostras correspondentes de dois dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem DLD-1 na ilha CpG a 5' do locus <i>MGC16121</i>	62
Figura 24. Padrão de metilação representativo de amostras correspondentes de dois dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem HCC1954 na ilha CpG a 5' do locus <i>MGC16121</i>	63
Figura 25. Padrão de metilação representativo de amostras correspondentes de um dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem DLD-1 na ilha CpG a 5' do locus <i>MGC16121</i>	64
Figura 26. Padrão de metilação representativo de amostras correspondentes de um dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem HCC1954 na ilha CpG a 5' do locus <i>MGC16121</i>	64
Figura 27. Padrão de metilação representativo de amostras correspondentes de um dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem DLD-1 na região mais a 5' da ilha CpG a 5' do locus <i>C596471</i>	65
Figura 28. Padrão de metilação representativo de amostras correspondentes de um dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem DLD-1 na região mais a 3' da ilha CpG a 5' do locus <i>C596471</i>	66
Figura 29. Padrão de metilação representativo de amostras correspondentes de dois dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem DLD-1 na região mais a 3' da ilha CpG a 5' do locus <i>C596471</i>	66

Figura 30. Padrão de metilação de cada dinucleotídeo CpG da ilha CpG posicionada na regiões 5' do locus <i>MGC16121</i>	69
Figura 31. Padrão de metilação de cada dinucleotídeo CpG da ilha CpG posicionada na região 5' do locus <i>CR596471</i> para a linhagem DLD1.....	70
Figura 32. Sequência consenso derivada da técnica RACE para extremidade 5' e representação gráfica do alinhamento da sequência consenso da região 5' da isoforma de <i>CR596471</i> que contém 3 éxons.....	71
Figura 33. Sequência consenso derivada da técnica RACE para extremidade 5' e representação gráfica do alinhamento da sequência consenso da região 5' da isoforma de <i>CR596471</i> que contém 2 éxons.....	72
Figura 34. Elemento repetitivo obtido quando utilizado o <i>primer Forward GSP</i> complementar à região do éxon 1 do locus <i>CR596471</i>	72
Figura 35. Sequências consenso obtidas a partir da técnica de RACE para a determinação das extremidades 3'	73
Figura 36. Sequência de 3 éxons com extremidade 3' mais curta de <i>CR596471</i>	74
Figura 37. Sequência de 3 éxons com extremidade 3' mais longa de <i>CR596471</i>	74
Figura 38. Sequência de 2 éxons com extremidade 3' mais curta de <i>CR596471</i>	75
Figura 39. Sequência de 2 éxons com extremidade 3' mais longa de <i>CR596471</i>	75
Figura 40. Sequência consenso derivada da técnica RACE para extremidade 3' e figura correspondente ao alinhamento para o gene <i>MGC16121</i>	76
Figura 41. Sequência consenso derivada da técnica RACE para extremidade 5' e figura correspondente ao alinhamento para o gene <i>MGC16121</i>	77
Figura 42. Sequência consenso do transcrito de <i>MGC16121</i>	78
Figura 43. Comparação entre isoformas do <i>CR596471</i> em relação ao perfil de expressão em painel de tecidos normais por meio de qRT-PCR.....	79
Figura 44. Comparação entre isoformas do <i>CR596471</i> em relação ao perfil de expressão em linhagens tumorais por meio de qRT-PCR.....	79
Figura 45. Correlação de Pearson (r) entre as isoformas de <i>CR596471</i> em relação às expressões obtidas das amostras de tecidos normais (A) e linhagens tumorais (B).....	80
Figura 46. Sequências de RNA semelhantes àquela encontrada para <i>CR595471</i> (sequência contendo 3 éxons) e seus detalhes.....	81
Figura 47. Sequências de RNA semelhantes àquela encontrada para <i>CR595471</i> (sequência contendo 2 éxons) e seus detalhes.....	81
Figura 48. Sequências das isoformas de <i>CR596471</i> constituídas de 2 éxons (Homo2 na figura) e 3 éxons (Homo3 na figura) alinhadas pela ferramenta <i>BLAT</i> ao genoma	

humano com as sequências de <i>Pan paniscus</i> (paniscus), <i>Pan troglodytes</i> (troglodytes) e <i>Gorilla gorilla</i> (Gorilla) obtidas no NCBI-BLAST.....	82
Figura 49. Estrutura secundária com menor energia livre do transcrito semelhante a <i>CR596471</i> obtido em <i>Pan paniscus</i>	82
Figura 50. Estrutura secundária com menor energia livre do transcrito semelhante a <i>CR596471</i> obtido em <i>Pan troglodytes</i>	83
Figura 51. Estrutura secundária com menor energia livre do transcrito semelhante a <i>CR596471</i> obtido em <i>Gorilla gorilla</i>	84
Figura 52. Estrutura secundária com do transcrito <i>CR596471</i> constituído de 3 éxons obtido em <i>Homo sapiens</i>	85
Figura 53. Estrutura secundária com menor energia livre do transcrito <i>CR596471</i> constituído de 2 éxons obtido em <i>Homo sapiens</i>	86
Figura 54. Sequências de RNA semelhantes àquela encontrada para <i>MGC16121</i> e seus detalhes.....	87
Figura 55. Sequência da isoforma obtida de <i>MGC16121</i> (Homo na figura) alinhada pela ferramenta <i>BLAT</i> ao genoma humano com as sequências de <i>Nomascus leucogenys</i> (Nomascus), <i>Callithrix jacchus</i> (Callithrix), <i>Papio anubis</i> (Papio) e <i>Saimiri boliviensis</i> (Saimiri) obtidas no NCBI-BLAST.....	87
Figura 56. Estrutura secundária com menor energia livre do transcrito de <i>MGC16121</i> obtido em <i>Homo sapiens</i>	88
Figura 57. Estrutura secundária com menor energia livre do transcrito semelhante a <i>MGC16121</i> obtido em <i>Nomascus leucogenys</i>	89
Figura 58. Estrutura secundária com menor energia livre do transcrito semelhante a <i>MGC16121</i> obtido em <i>Callithrix jacchus</i>	90
Figura 59. Estrutura secundária com menor energia livre do transcrito semelhante a <i>MGC16121</i> obtido em <i>Saimiri boliviensis</i>	91
Figura 60. Estrutura secundária com menor energia livre do transcrito semelhante a <i>MGC16121</i> obtido em <i>Papio anubis</i>	92
Figura 61. Parte da sequência do transcrito de <i>MGC16121</i> , correspondente à subestrutura secundária encontrada também nos transcritos de <i>Nomascus leucogenys</i> , <i>Callithrix jacchus</i> e <i>Saimiri boliviensis</i>	92

SUMÁRIO

1.Introdução.....	20
1.1. Placenta e células germinativas X câncer.....	20
1.2. Regulação da expressão gênica	23
1.2.1. Metilação do DNA	23
1.2.2. RNAs não codificadores.....	25
1.2.2.1. microRNAs.....	26
1.2.2.2. RNAs longos não codificadores.....	27
1.3. Cromossomo X e Região Xq26.....	30
1.4. Genes <i>MGC16121</i> e <i>CR596471</i>	32
2. Justificativa.....	35
3. Objetivos.....	36
3.1. Objetivo geral	36
3.2. Objetivos específicos	36
4. Materiais e métodos.....	37
4.1. Manutenção e expansão das linhagens celulares	38
4.2. Extração de RNA (Figuras 3 A.2 e 3B.2)	38
4.3. Síntese de cDNA (Figuras 3 A.2a e 3B.2)	38
4.4. Análise dos níveis de expressão através de qRT-PCR (Figuras 3 A.2a e 3 B.2a)	39
4.4.1. Análise de expressão dos genes <i>MGC16121</i> e <i>CR596471</i> e miRNAs da vizinhança em amostras de placenta (Figuras 3 A.1)	39
4.5. RACE (<i>Rapid Amplification of cDNA Ends</i>) (Figura 3 A2.b)	40
4.5.1 Análise dos transformantes	42
4.6. Estudo de metilação (Figura 3B)	43
4.7. Extração de DNA e PCR após o tratamento com o 5-aza-2-deoxicitidina (Figura 3 B.3)	44
4.8. Modificação do DNA por bissulfito de sódio (Figura 3 B.4)	44
4.9. Análise dos níveis de metilação por MS-HRM (<i>Methylation-Sensitive High Resolution Melting</i>) (Figura 3 B.4a)	44
4.11. Análise da conservação e estrutura secundária dos transcritos.....	47
4.12. Forma de análise dos resultados	48
5. Resultados.....	49
5.1. Perfil de expressão dos genes <i>MGC16121</i> e <i>CR596471</i> , e miRNAs da vizinhança em um painel de tecidos normais.....	49

5.2. Perfil de expressão dos genes <i>MGC16121</i> e <i>CR596471</i> , e miRNAs da vizinhança em linhagens tumorais	51
5.3. Comparação entre os níveis de expressão dos genes <i>MGC16121</i> , <i>CR596471</i> e miRNAs da vizinhança em tecidos normais e linhagens tumorais.....	54
5.4. Análise de expressão dos genes <i>MGC16121</i> e <i>CR596471</i> , e miRNAs da vizinhança em amostras de placenta.....	55
5.5. Análise dos níveis de expressão após o tratamento com 5-aza-2-deoxicitidina.....	56
5.6. Análise dos níveis de metilação através de MS-HRM (<i>Methylation-Sensitive High Resolution Melting</i>).....	62
5.7. Caracterização do padrão de metilação dos promotores dos genes <i>MGC16121</i> e <i>CR596471</i> pelo sequenciamento das ilhas CpG	67
5.8. Caracterização das extremidades 5' e 3' dos transcritos dos genes <i>CR596471</i> e <i>MGC16121</i>	71
5.9. Comparação dos perfis de expressão entre as isoformas do gene <i>CR596471</i> em um painel de tecidos normais e linhagens tumorais	78
5.10. Análise da conservação e estrutura secundária dos transcritos de <i>CR596471</i> e <i>MGC16121</i>	80
6. Discussão.....	93
6.1. Análise de expressão e correção.....	93
6.1.1. Análise de expressão das isoformas do gene <i>CR596471</i>	95
6.1.2. Análise de expressão dos genes <i>MGC16121</i> e <i>CR596471</i> , e miRNAs da vizinhança em amostras de placenta.....	96
6.2. Regulação da expressão por metilação.....	96
6.3. Estrutura e conservação	97
7. Conclusões.....	100
8. Referências bibliográficas.....	102
9. Apêndice.....	111
10. Anexos.....	112

1. INTRODUÇÃO

1.1. Placenta e células germinativas X câncer

A placenta é um órgão extra-embriônico complexo cuja função consiste em facilitar a troca de nutrientes e gases entre a mãe e o feto, além de protegê-lo do sistema imunológico materno (Rawn & Cross, 2008). As células germinativas são células diplóides que dão origem aos gametas haplóides, células especializadas que eventualmente, quando há fertilização, podem gerar um novo organismo diplóide.

A despeito de suas funções, as células germinativas e a placenta, como células/tecidos normais, apresentam características semelhantes àsquelas encontradas em células malignas (Simpson *et al.*, 2005; Ferretti *et al.*, 2007;).

Foi o embriologista John Beard, em 1902, quem primeiramente observou similaridades entre as células trofoblásticas, que darão origem à placenta, e as células tumorais em sua “Teoria Trofoblástica do Câncer”. Ele propôs que as últimas seriam derivadas das células germinativas primárias errantes. Tais células que falhassem a completar sua migração embrionária para as gônadas e então viessem para locais diferentes de seu destino, formariam tumores semelhantes fenotipicamente às células trofoblásticas (Simpson *et al.*, 2005). Apesar de esta teoria ser incorreta quanto à célula de origem tumoral, de fato, a placenta seria semelhante a um “tumor bem comportado” (Soundararajan & Rao, 2004).

Muitas características importantes do fenômeno da carcinogênese são compartilhadas ou correspondentes ao processo da placentação e gametogênese (Figura 1). As características fenotípicas correspondentes entre células tumorais e as células germinativas ou trofoblastos incluem a imortalização (envolvida na transformação), invasão, a indução da meiose (correspondente à aneuploidia em células tumorais) e a migração (que contribui para a metástase). Já os fenótipos compartilhados entre esses tipos celulares incluem a demetilação, a indução da angiogênese, regulação do complexo maior de histocompatibilidade (evasão imunológica) e a expressão da gonadotrofina coriônica (Figura 1) (Simpson *et al.*, 2005). Apesar disso, a placenta, mais especificamente, pode ser vista como a contraparte fisiológica de tumores

malignos, visto que esta para de crescer em determinado momento, não se metastizando (Soundararajan & Rao, 2004).

Como pode ser predito a partir dos comportamentos semelhantes supracitados, a placenta também apresenta circuitos moleculares compartilhados com células tumorais. Genes expressos de maneira similar incluem proto-oncogenes; fatores de crescimento e seus receptores; enzimas; hormônios e antígenos associados a tumores (Ferretti *et al.*, 2007). Já as células germinativas começaram a ser associadas a células tumorais a partir da observação de que vários tumores também produzem o hormônio gonadotrofina coriônica (Akers *et al.*, 2010). Depois dessa descoberta, várias outras proteínas que são expressas somente em células germinativas e placenta, mas que são expressas de forma anormal em diferentes tipos de cânceres foram relatadas (Junbluth *et al.*, 2007; Caballero & Chen, 2009). Elas são conhecidas como “antígenos câncer-testículo” (antígenos CT) e são muito estudadas pelo seu potencial terapêutico em vacinas anti-cânceres (Parmigiani *et al.*, 2006).

Todos estes fatos supracitados formam a base da teoria de que genes de células germinativas e trofoblásticas, que são expressos de forma anormal em células tumorais “...refletem a reativação do programa gametogênico silenciado em células somáticas, como uma das forças que dirige a tumorigênese” (Simpson *et al.*, 2005). Dessa forma, podemos dizer que o câncer não é só uma doença do crescimento, mas também da reprodução. Portanto, com base na teoria trofoblástica do câncer, a recapitulação da gametogênese pela célula tumoral seria correspondente a uma “gravidez de célula somática” (modificado de Old, 2007).

Pelo exposto neste item, o estudo dos genes expressos na placenta e células germinativas, pode ser considerado como um sistema modelo para investigação dos processos celulares da tumorigênese (modificado de Soundararajan & Rao, 2004).

Fenótipos correspondentes		Fenótipos compartilhados	
Gametogênese	Tumorigênese		
Imortalização	2	Transformação	1
Implantação	4	Invasão	3
Meiose	5	Aneuploidia	7
Migração	6	Metástase	8
			9
			Expressão de gonadotrofina coriônica

Figura 1. Semelhanças entre células germinativas/trofoblastos e células tumorais. A ativação do programa gametogênico (células marrons) poderia contribuir para as propriedades da formação e progressão tumoral (células azuis). Os números (1-9) indicam características fenotípicas e os estágios nos quais estes eventos ocorrem. (Figura e legenda modificados de Simpson *et al.*, 2005).

1.2. Regulação da expressão gênica

Todas as células de um organismo possuem, com algumas exceções, conteúdo genômico idêntico. Os fenótipos diferentes observados entre elas são derivados de mecanismos regulatórios que independem da sequência de DNA e são herdados ao longo das divisões celulares (Jurkowska *et al.*, 2011). O estudo dessas alterações herdáveis é denominado epigenética (Tollefsbol, 2009). A epigenética abrange a metilação do DNA, os RNA não codificadores e as modificações covalentes de proteínas histona. Os dois primeiros são abordados no presente trabalho.

1.2.1. Metilação do DNA

A metilação do DNA, a adição de radicais metil no DNA que leva ao silenciamento gênico, é o mecanismo epigenético de regulação gênica mais estudado (Tollefsbol, 2009). Ela foi revelada pela primeira vez em 1948 por Hotchkiss em amostra de timo de bezerro utilizando cromatografia de papel (Tollefsbol, 2009).

A maioria dos radicais metil (CH_3) são encontrados ligados ao carbono 5 de citosinas pertencentes ao dinucleotídeo CpG, sequências 5'-CG-3', as quais são metiladas pela ação das DNA-metiltransferases DNMT1, DNMT3a e DNMT3b (Robertson, 2001). Em humanos, cerca de 70% dos dinucleotídeos CpG são metilados (Kondo, 2009). A DNMT1 realiza a manutenção da metilação após cada duplicação do DNA, tornando possível sua herdabilidade, e as DNMT3a e DNMT3b, a metilação *de novo* em resposta a sinais ambientais (Mazzio & Soliman, 2012).

Embora a quantidade de nucleotídeos CpG encontrada no genoma humano seja menor do que a esperada, correspondendo a apenas 2-5% (Lopez *et al.*, 2009), há regiões ricas em CpG (entre 0,5 e 2 kb de extensão) cuja proporção relativa de G + C é maior em relação ao restante do DNA. Estas regiões recebem o nome de ilhas CpG (Bird, 1986). As ilhas CpG encontram-se geralmente na região promotora próxima ao sítio de transcrição de genes constitutivos em regiões não metiladas (Lopez *et al.*, 2009). Já as regiões metiladas do genoma consistem em sequências repetitivas que incluem transposons e retrovírus endógenos cuja inativação transcricional está relacionada à proteção contra a proliferação desses elementos pelo DNA (Bestor, 1998).

As citosinas dos dinucleotídeos CpGs das regiões reguladoras, quando metiladas, alteram a estrutura da cromatina e regulam negativamente a expressão gênica, impedindo que fatores de transcrição se liguem à sítios de ligação ou que reguladores especializados que se ligam de maneira específica ao DNA metilado (Buck-Koehntop & Defossez, 2013). As proteínas que interpretam o sinal de metilação do DNA são denominadas proteínas de ligação aos grupos metil-CpG (*Metil-CpG Binding Proteins* - MBPs) (Tollefsbol, 2009). Em humanos, já foram encontradas 15 dessas proteínas, que se dividem em 3 grupos estruturais: as família MBD, a família *zinc finger* e a SRA (Parry & Clarke, 2011; Buck-Koehntop & Defossez, 2013). Cada grupo utiliza um mecanismo diferente para interagir com o DNA metilado e seus componentes (Parry & Clarke, 2011).

A enzima DNMT1 liga-se diretamente a MBPs, que também se associam a enzimas constritivas modificadoras de histonas (Mazzio & Soliman, 2012). Essas enzimas modificam covalentemente as histonas, causando uma maior compactação da cromatina por meio de alterações eletrostáticas entre o DNA e as histonas da região afetada (modificado de Wilson, 2008). Tais enzimas constritivas incluem as histonas deacetilases (HDAC1 e HDAC2), a histona metiltransferase (HMT) e a proteína heterocromatina 1 (HP1). Este complexo modificador do DNA e cromatina é um dos responsáveis pelo silenciamento gênico, (Feinberg & Tycko, 2004; modificado de Mazzio & Soliman, 2012). Portanto, a metilação do DNA e as modificações covalentes das histonas atuam sinergicamente, deixando a cromatina compactada e, conseqüentemente, menos acessível à ligação de fatores de transcrição (Daniel *et al.*, 2010).

A metilação do DNA está envolvida não só em mecanismos normais, como *imprinting* gênico, inativação do cromossomo X, desenvolvimento do sistema imunológico, reprogramação celular e comportamento, mas também no desenvolvimento de várias doenças humanas, como as doenças psiquiátricas, as do sistema imunológico e do câncer (Jurkowska *et al.*, 2011).

Nas células tumorais, o padrão de metilação é alterado e consiste de eventos de hipometilação global do genoma e hipermetilação das regiões reguladoras próximas aos genes (Robertson, 2001).

A hipometilação global, ou seja, a desmetilação do DNA em regiões não promotoras, pode causar instabilidade genômica e mudanças estruturais no genoma. De modo mais específico, a desmetilação dessas regiões pode levar à expressão nociva de

transposons e retrovírus endógenos normalmente reprimidos e afetar a estabilidade funcional dos cromossomos, como nas regiões pericentroméricas. Por outro lado, a hipometilação também pode afetar promotores de oncogenes, levando ao aumento de sua expressão e contribuindo para o processo carcinogênico. Já a hipermetilação de ilhas CpGs associadas a promotores pode silenciar genes supressores tumorais (Daniel *et al.*, 2010).

Além disso, antígenos CT associados ao cromossomo X podem ter suas regiões promotoras desmetiladas, sendo ativados nas células tumorais. Este evento poderia mediar o fenótipo maligno na ausência de mutações em oncogenes e genes supressores tumorais conhecidos (Simpson *et al.*, 2005).

O padrão aberrante de metilação pode fornecer vantagem de crescimento à célula, de modo similar às mutações e deleções (Robertson, 2001). Consequentemente, o silenciamento epigenético é atualmente reconhecido como a “terceira via” no modelo de Knudson da inativação de genes supressores tumorais no câncer, podendo alterar a função do gene sem modificar a sequência do DNA (Kondo & Issa, 2004).

1.2.2. RNAs não codificadores

Contrariando todas as estimativas da época, o Projeto Genoma Humano contabilizou aproximadamente 25.000 genes codificadores de proteínas, quando o número esperado era maior do que 60.000 (Ponting & Belgard, 2010). O número encontrado de genes codificadores em humanos não é muito diferente daquele de organismos menos complexos, como o de *Caenorhabditis elegans*, sugerindo que os genes não codificadores possuiriam maior importância para o aumento da complexidade dos organismos eucariotos (Costa, 2010). Os genes codificadores ocupam somente 2% do genoma (Gibb *et al.*, 2011) e, por meio de novas tecnologias de sequenciamento como *tiling arrays*, análises do transcriptoma inteiro e RNA-Seq foi demonstrado atualmente que os genomas dos eucariotos são transcritos em mais de 90% de sua extensão (Costa, 2010).

Entre os transcritos não codificadores foram incluídos; além dos RNAs classicamente já conhecidos, como mRNAs (RNAs mensageiros), tRNAs (RNAs transportadores) e rRNAs (RNAs ribossômicos); os pequenos RNAs reguladores, como miRNAs (microRNAs), siRNAs (pequenos RNAs de interferência), piRNAs (RNAs que interagem com proteínas Piwi), snRNAs (pequenos RNA nucleares) e snoRNAs

(pequenos RNAs nucleolares) e os lncRNAs (RNAs longos não codificadores) (Wapinski & Chang, 2011).

Dentro do grupo dos RNAs são abordados aqui os miRNAs e os lncRNAs, categorias as quais pertencem os transcritos estudados neste trabalho.

1.2.2.1. microRNAs

Os microRNAs (miRNAs) são RNAs não codificadores de cerca de 22 pb que possuem papel regulatório importante na expressão gênica em variados processos biológicos, como no desenvolvimento; proliferação celular; diferenciação; metabolismo e apoptose (Bartel, 2004). Estes também estão relacionados a doenças como o câncer (Huntzinger & Izauralde, 2011). A função dos miRNAs consiste em silenciar genes através de clivagem ou repressão transcricional (Bartel, 2004).

Os miRNAs são transcritos pela RNA polimerase II como pri-miRNAs, estruturas em forma de haste-laço. Eles são identificados e processados pelo complexo microprocessador que contém a enzima RNase III Drosha e o cofator DGCR8. Tal complexo digere o pri-miRNA em uma estrutura menor denominada pre-miRNA. Este, por sua vez, é transportado do núcleo para o citoplasma pelo complexo constituído pela Exportina-5 e uma GTPase. No citoplasma o pre-miRNA é também processado por outra RNase III denominada Dicer associada à proteína TRBP, gerando um miRNA de fita dupla. Uma das fitas é degradada e a outra é associada ao complexo miRISC, que é composto pela proteínas Argonata, GW182, outras proteínas acessórias e o miRNA maduro. Este complexo reconhece a região 3' não traduzida (3' UTR) de um mRNA específico e de complementaridade perfeita ou imperfeita com a região 5' do miRNA. No caso de haver complementaridade imperfeita, a mais usual, a tradução do mRNA é bloqueada; já no caso em que a complementaridade é perfeita, o mRNA é degradado. (Hussain, 2012).

Dentre os miRNAs já estudados, 50% estão organizados em *clusters* e, quando sob controle de um único promotor, podem ser expressos como transcritos policistrônicos (Lages *et al.*, 2012). Os miRNAs de determinado *cluster* são considerados coexpressos por meio das evidências dos dados do perfil de expressão de vários tecidos e linhagens celulares. Isto significa que, em condições anormais decorrentes do estabelecimento de uma doença, caso haja desregulação de um miRNA

do *cluster*, os outros membros deste também deverão apresentar uma desregulação similar (Farazi *et al.*, 2011).

Atualmente está bem estabelecido que a desregulação de miRNAs pode dirigir ou contribuir para a tumorigênese. miRNAs superexpressos em tumores podem agir como oncogenes se seus alvos são genes supressores tumorais; já miRNAs subexpressos quando do estabelecimento do câncer podem agir como supressores tumorais se seus alvos são oncogenes (Croce, 2009; Farazi *et al.*, 2011).

A desregulação dos miRNAs na tumorigênese pode-se dar por meio de vários mecanismos, como defeitos em sua biogênese; rearranjos cromossômicos; mutações de ponto em genes codificadores ou devido à desregulação transcricional. Dentro deste último mecanismo estão incluídas alterações nas modificações epigenéticas nas regiões promotoras de genes de miRNAs, sendo a metilação do DNA uma das mais importantes destas alterações (Lages *et al.*, 2012). Dessa forma, miRNAs que agem como supressores tumorais podem ser silenciados através da hipermetilação de ilhas CpGs em sua regiões promotoras (Croce, 2009).

O controle da expressão de determinado miRNA por metilação do DNA pode ser confirmado por meio de sua superexpressão após o uso de drogas demetilantes, como a 5-Aza-2'-deoxicitidina, que inibe a DNA metiltransferase (DNMT), em culturas de células tumorais (Croce, 2009; Lages *et al.*, 2012).

Os perfis de expressão de miRNAs podem diferenciar tecidos tumorais de sua contraparte normal, podem discriminar diferentes tipos de tumores e indicar o tecido de origem do tumor quando não é possível obter esta informação por outros meios, como pela histologia e informações clínicas. Dessa maneira, os miRNAs são potenciais biomarcadores clínicos para o diagnóstico e prognóstico de pacientes (Croce, 2009; Farazi *et al.*, 2011).

Os miRNAs são altamente conservados mesmo entre organismos evolutivamente distantes e apesar de só constituírem 1 a 3% do genoma humano, controlam mais de 30% dos seus genes (Hussain, 2012). Os miRNAs estão presentes em todos os cromossomos, com exceção do Y (Lages *et al.*, 2012). 10% dos miRNAs humanos atualmente conhecidos estão localizados no cromossomo X, sendo que vários destes possuem participação no estabelecimento e progressão do câncer e regulam o sistema imunológico (Pinheiro *et al.*, 2011).

1.2.2.2. RNAs longos não codificadores

RNAs longos não codificadores (lncRNAs) são transcritos semelhantes aos mRNAs, pois como estes apresentam extremidades 5' com adição de 7-metilguanosina e extremidades 3' poliadeniladas, podem sofrer *splicing* e são transcritos pela RNA polimerase II. No entanto não apresentam uma matriz aberta de leitura longa (ORF – *Open Reading Frame*) suficiente para codificar proteínas, sendo que seu tamanho varia entre 200 a 100.000 pares de bases (Moran *et al.*, 2012).

Os lncRNAs já haviam sido estudados antes mesmo da descoberta dos miRNAs, porém ainda não possuíam esta classificação. Os primeiros lncRNAs a serem estudados foram o gene imprintado *H19* e o Transcrito Específico do X Inativo (*XIST*) no começo da década de 1990, ambos envolvidos em mecanismos epigenéticos. Todavia, houve mudança de foco da pesquisa sobre RNAs não codificadores, dos longos para os pequenos, com o relato do primeiro miRNA (lin-14) em 1993 (Gibb *et al.*, 2011). Com a publicação recente de que o genoma é transcrito de forma ubíqua, sendo que a maioria de suas bases podem ser encontradas em transcritos primários não codificadores e a identificação de muitos destes (The ENCODE Project Consortium, 2007), houve aumento da pesquisa visando à descoberta e caracterização de novos lncRNAs (Gibb *et al.*, 2011), os RNAs menos conhecidos até o momento.

Apesar de serem menos conservados do que genes codificadores em relação à sequência de nucleotídeos, os lncRNAs apresentam uma maior conservação evolutiva de suas estruturas secundárias (Torarinsson *et al.*, 2006). Dessa forma, os lncRNAs não são mais vistos como somente “ruído” transcricional (Mitra *et al.*, 2012), sendo recentemente considerados como reguladores chave de diversos processos biológicos (Liu *et al.*, 2012).

Os lncRNAs podem ser transcritos a partir de regiões distantes dos genes codificadores, dentro de tais genes ou transcritos a partir de íntros. Eles podem exercer sua ação em *cis* ou *trans* a partir da região de origem, regulando seus alvos, genes codificadores ou não, de forma positiva ou negativa (Wapinski & Chang, 2011).

Os lncRNAs derivados da fita de DNA oposta de um gene codificador são conhecidos como transcritos antisenso naturais (NATs) e regulam em *cis* o gene o qual se sobrepõem (Moran *et al.*, 2012). Estes lncRNAs incluem os pseudogenes, cópias de determinado gene que não produzem uma proteína funcional (Hirotsume *et al.*, 2003). Como exemplo de um pseudogene que possui uma função já bem estudada pode-se citar o par gene-pseudogene *PTEN-PTENP1* (*Phosphatase and tensin homolog - Phosphatase and Tensin Homolog Pseudogene 1*). O pseudogene *PTENP1* produz um

transcrito com alta homologia à região 3' UTR do transcrito do gene supressor tumoral *PTEN*. Assim sendo, *PTENP1* compete pelos miRNAs que alvejam o transcrito de *PTEN* e, como resultado, a superexpressão de *PTENP1* faz com que aumente os níveis da proteína PTEN (Muro *et al.*, 2011).

A maioria dos lncRNAs com função caracterizada até o momento são reguladores epigenéticos (modificado de Costa, 2010). Estes lncRNAs se associam a complexos remodeladores de cromatina, grupo de genes que reestruturam os nucleossomos de modo a deixar a cromatina mais ou menos compactada, determinando o nível de transcrição gênica de uma região definida do cromossomo. Exemplos destes lncRNAs incluem o *Xist*, que é responsável pelo silenciamento do cromossomo X a ser inativado em fêmeas de alguns animais como mecanismo de compensação de dose; o *Airn* e *Kcnqlot1* responsáveis pelo *imprinting* paterno-específico de alelos localizados nos *clusters* do *Igfr2* e *Kcnql1*, respectivamente; e o *HOTAIR*, que determina o silenciamento de alguns genes do *cluster HOXD* (Saxena & Carninci, 2011).

Os complexos remodeladores de cromatina como PRC1, PRC2, G9a e LSD1 contêm enzimas que possuem domínios de ligação a RNAs, sendo que cada complexo pode se ligar a diferentes lncRNAs e estes, a diferentes complexos. As interações entre proteínas do complexo e lncRNAs poderiam resultar em mudanças conformacionais que seriam úteis para distinguir a especificidade da região alvo (Saxena & Carninci, 2011).

Em resumo, tais lncRNAs serviriam como “guias” para os complexos remodeladores de cromatina, pois estes não possuem capacidade de ligação ao DNA, não reconhecendo suas regiões alvo de forma isolada (Moran *et al.*, 2012). Outras funções dos lncRNAs estão relacionadas ao processamento de mRNA (*splicing*) e ao aumento ou diminuição da capacidade de tradução de seus alvos (Wapinski & Chang, 2011).

À luz das novas descobertas que revelaram a transcrição onipresente do genoma e o envolvimento global dos lncRNAs em sua regulação, não foi difícil associar a desregulação de lncRNAs à doenças. Dentre essas se pode citar as doenças metabólicas; neurodegenerativas; psiquiátricas; cardiovasculares; a auto-imunidade; o mal de Alzheimer; a Síndrome do X frágil e o câncer (Wapinski & Chang, 2011; Moran *et al.*, 2012).

1.2.2.2.1. RNAs longos não codificadores e câncer

A expressão diferencial de diversos lncRNAs em tecidos normais *versus* tumorais foi a primeira evidência de que a desregulação de lncRNAs poderia não ser somente uma consequência do processo tumorigênico, mas sim uma de suas causas, uma vez que eles controlam a transcrição de várias regiões do genoma (Tsai *et al.*, 2011). Variados tipos de cânceres já foram associados à desregulação de lncRNAs, como câncer de mama; de próstata; carcinoma hepatocelular; leucemia e melanoma (Moran *et al.*, 2012), o que valida os lncRNAs como elementos importantes no processo de tumorigênese em geral e não só associado a um único tipo de tecido.

Alguns dos lncRNAs já descritos como tendo um papel na carcinogênese são: o *SPRY4-IT1*, superexpresso em melanomas (Yan & Wang, 2012); o RNA antisenso p15 e o lincP21, que agem como supressores tumorais (Moran *et al.*, 2012); os já citados *H19*, altamente expresso em metástases de fígado derivadas de vários tipos de tumores primários, e o *HOTAIR*, cuja superexpressão foi associada ao carcinoma hepatocelular, câncer colorretal e de mama (Wapinski & Chang, 2011).

Tendo em vista a expressão aberrante de lncRNAs em tecidos tumorais, surgem novas possibilidades de que estes possam ser utilizados como biomarcadores de diagnósticos e prognósticos, além de poderem ser candidatos a alvos de drogas em novas terapias (modificado de Moran *et al.*, 2012). Sendo assim, a caracterização de novos lncRNAs, além de ser um campo novo de estudo, é uma necessidade real e importante para o desenvolvimento desses novos tratamentos para cânceres (Tsai *et al.*, 2011; modificado de Wapinski & Chang, 2011).

1.3. Cromossomo X e Região Xq26

Há evidências de que o cromossomo X contenha maior proporção de genes relacionados ao sexo, reprodução e funções cerebrais do que seria esperado de forma aleatória (Hurst & Randerson, 1999). De fato, no cromossomo X há pelo menos três vezes mais genes que são relacionados ao sexo e reprodução, e cinco vezes mais genes que são expressos no cérebro em comparação aos outros cromossomos humanos (Graves, 2010).

Tal disparidade poderia ser explicada pela idéia de que “os cromossomos sexuais têm um importante papel na evolução de traços sexuais dimórficos (quando machos e

fêmeas possuem estruturas ou formas diferentes)” (Rice, 1984). Em sua teoria, Rice explica que características sexuais masculinas vantajosas como chifres maiores seriam acumuladas no cromossomo X. Isso porque alelos recentemente adquiridos que atribuíssem tais vantagens aos machos seriam mais facilmente selecionados pelas fêmeas por estarem expostos, sendo recessivos ou dominantes, em indivíduos XY (Graves, 2010). Dessa forma, o cromossomo X teria acumulado genes relacionados ao sexo, desenvolvimento e reprodução e também aqueles relacionados a funções cerebrais. Estes são, de maneira geral, genes que controlam características que seriam selecionadas pelas fêmeas, como o tamanho corporal; tamanho de cérebro; quantidade de esperma; tamanho de chifres, etc.

Como resultado da seleção de genes vantajosos para machos no cromossomo X, este acumulou os antígenos CT (Graves, 2010). Os antígenos CT representam 10% do total de genes do cromossomo X, sendo a região entre Xq24-q28 aquela que contém a maior densidade de genes CT e incluem as famílias multigênicas MAGE-A, MAGE-C e NY-ESO-1 (Fratta et al., 2011). Mais especificamente, acredita-se que a região Xq26 seja importante para o desenvolvimento normal da placenta e do feto (Fant *et al.*, 2010). O gene CT *PLAC1* está localizado nesta região e sua descrição é importante para o entendimento da proposta do presente trabalho.

O gene *PLAC1*, relatado pela primeira vez por Cocchia *et al.*, (2000) possui função ainda não esclarecida e sua expressão é limitada à placenta. A proteína codificada por este gene é regulada pelos genes de crescimento *FGF-7* e *EGF* durante a diferenciação do trofoblasto, considerado a primeira linhagem celular a se diferenciar e dar origem a maioria dos tecidos extra-embrionários (Massabbal *et al.*, 2005). Além disso, baixas concentrações de seu mRNA no sangue materno estão relacionadas ao risco de aborto durante a gravidez (Farina *et al.*, 2005). A proteína PLAC1 está presente em vários tipos de tumores e envolvida nos processos celulares que tornam as células cancerosas malignas (Koslowski *et al.*, 2007).

O gene *PLAC1* está localizado a cerca de 65 kb do gene *HPRT1* (UCSC Genome Bioinformatics), que é responsável pela produção da enzima hipoxantina guanina fosforibosiltransferase cuja deficiência leva a síndrome de Lesch-Nyhan em humanos, caracterizada pela super produção do ácido úrico, automutilação, espasticidade, corioaterose e retardo mental variável (Nussbaum *et al.*, 2001). Porém, a deleção do gene homólogo em embriões de camundongos derivados de células-tronco embrionárias, sem atividade da enzima, não levou ao fenótipo observado em humanos

(Searle *et al.*, 1994). Apesar deste resultado, em 1998, quando Kushi e colaboradores produziram camundongos quimeras de linhagem germinativa com deleções de 200 a 700 kb ao redor do locus *Hprt*, que apresentavam um terço do tamanho dos animais controles ou eram natimortos. Logo, a causa mais aceitável para a produção de fenótipos aberrantes são defeitos em outros genes que flanqueiam o locus do *Hprt*.

Até recentemente o gene *PLAC1* era o único candidato na região Xq26 associado com o desenvolvimento placentário normal. Todavia, na última versão do Genome Browse – UCSC (Feb. 2009, CRCh37/hg19), outros genes foram adicionados a essa região: *PHF6*, *MGC16121* e *CR596471* (Fant *et al.*, 2010).

1.4. Genes *MGC16121* e *CR596471*

Dos três genes citados no item anterior, os genes *MGC16121* e *CR596471* ainda não haviam sido caracterizados quanto à estrutura e função. Ambos estão situados entre os loci do *HPRT1* e *PLAC1*, em orientações opostas e distantes cerca de 3 kb. Cada um apresenta três éxons e ilhas CpG posicionadas na região inicial dos genes. É observado a presença de vários microRNAs (miRNAs) situados próximos dos genes *MGC16121* e *CR596471*. Há forte evidência de que os miRNAs miR-503 e miR-424 sejam hospedeiros do gene *MGC16121* (Figura 2) (Genome Browse – UCSC; Feb. 2009, CRCh37/hg19).

De acordo com o Genome Browse – UCSC (Feb. 2009, CRCh37/hg19), tanto o *CR596471* quanto o *MGC16121* são descritos como transcritos não-codificadores (ncRNAs), ou seja, seus RNAs não são traduzidos em proteínas, e pelos seus tamanhos (maiores de 200 pares de bases), podem ser considerados RNAs longos não codificadores (lncRNAs).

Genes silenciados na maioria dos tecidos saudáveis, mas expressos em tecidos ou linhagens gametogênicas e/ou trofoblásticas são frequentemente ativados e expressos ectopicamente em cânceres humanos. Tal característica é importante, já que a identificação de genes tumores-específicos é um dos pré-requisitos para o desenvolvimento de novos métodos de diagnóstico e terapias contra cânceres (modificado de Koslowski *et al.*, 2007).

Resultados preliminares produzidos por nosso laboratório indicam que o gene *MGC16121* tem expressão restrita à placenta e apresenta evidência de ser regulado por

mecanismos de metilação na linhagem MCF-7, derivada de adenocarcinoma de mama. Já o *CR596471* está expresso em alguns tipos de tecidos normais, como a placenta, próstata, ovário, testículo e coração. No entanto, sua expressão é maior na placenta que nos demais tecidos. Essas observações sugerem que tais loci possam desempenhar papéis semelhantes ao do *PLAC1*.

No presente trabalho, também foram analisados os miRNAs na região Xq26 próximos aos genes *MGC16121* e *CR596471* (miR-424, 503, 450a, 450b-5p e 542-3p). Esta análise possuía a intenção de verificar possíveis correlações entre eles, como forma de evidenciar a possibilidade de coexpressão entre seus membros e os genes do estudo tanto em condições normais, como em tumorais.

Figura 2. Localização dos genes *MGC16121* e *CR596471*. A. Localização em relação aos loci *HPRT1* e *PLAC1*. B. Região ampliada da figura A mostrando a estrutura dos genes *MGC16121* e *CR596471* (recentemente modificado para *LOC100506757*), as ilhas CpG que os intercepta e os miRNAs. As sequências de interesse estão dentro do retângulo em cor vermelha. Modificado de Genome Browse – UCSC (versão: Feb. 2009, CRCh37/hg19).

2. JUSTIFICATIVA

Como descrito anteriormente, os genes *MGC16121* e *CR596471* estão localizados em uma região cromossômica que concentra genes associados com a reprodução e desenvolvimento. Considerando que o padrão de expressão de ambos parece ser restrito a poucos tecidos normais (placenta, testículo e ovário), sendo mais expressos em placenta e em algumas linhagens tumorais, similar ao padrão de expressão do gene *PLAC1*, que é um antígeno tumoral testado como imunoterápico em diversos tumores. Portanto, os dois genes passam a ser potenciais candidatos a marcadores tumorais. Além disso, pelo fato de que um deles possa ser hospedeiro de dois miRNAs (miR-424 e miR-503) e a evidência de que pelo menos um deles seja regulado pelo mecanismo de metilação, o presente estudo tem como objetivo central a caracterização da estrutura e regulação dos genes *MGC16121* e *CR596471*.

3. OBJETIVOS

3.1. Objetivo geral

Caracterizar a estrutura, a expressão tecidual e o mecanismo de regulação por metilação dos genes *MGC16121* e *CR596471*.

3.2. Objetivos específicos

- Caracterizar a estrutura dos transcritos dos genes *MGC16121* e *CR596471*;
- Comparar os níveis de expressão dos genes *MGC16121* e *CR596471* e miRNAs miR-424, 503, 450a, 450b-5p e 542-3p em linhagens tumorais, em um painel de tecidos normais e em amostras de placenta;
- Comparar os níveis de expressão dos genes *MGC16121* e *CR596471* e miRNAs miR-424, 503, 450a, 450b-5p e 542-3p em algumas linhagens tratadas e não tratadas com um agente demetilante;
- Comparar o padrão de metilação das ilhas CpG dos genes *MGC16121* e *CR596471* em linhagens tratadas e não tratadas com o agente demetilante.

4. MATERIAIS E MÉTODOS

O presente projeto foi aprovado pelo comitê de ética do HC-FMRPUSP, processo nº 13867/2011, (documento de aprovação contida no anexo A) e realizado conforme o esquema da figura 3.

Figura 3. Representação esquemática das atividades laboratoriais que foram realizadas no presente projeto. A. Análise de expressão dos genes *MGC16121* e *CR596471*, e miRNAs das proximidades; e definição das extremidades 5' e 3' dos genes do estudo. B. Análise da expressão, níveis e padrão de metilação em linhagens tratadas e não tratadas com o agente demetilante para os genes citados e miRNAs da vizinhança.

4.1. Manutenção e expansão das linhagens celulares

Nosso laboratório possui um banco de linhagens o qual foi utilizado para a análise do perfil de expressão. Linhagens de carcinoma de células escamosas de cabeça e pescoço (FaDu, UM-SCC-14 e UM-SCC-38); adenocarcinoma colorretal (DLD-1 e HT-29); carcinoma hepatocelular (HepG2); leucemia (Jurkat, K562 e NB4); linfoma (U937); tumor de mama (HCC1954, MCF-7 e SKBR3); coriocarcinoma (JEG-3); carcinoma de próstata (LnCap); melanoma (SK-MEL-37); glioblastoma (T98G) e a linhagem tumorigênica HEK293, derivada de rim embrionário, foram cultivadas por pelo menos mais de uma passagem após o descongelamento de alíquotas em meios de cultivo apropriados preparados conforme especificações comerciais. As linhagens foram mantidas em estufa a 37°C sob 5% de CO₂.

Além do RNA proveniente de linhagens de origem tumoral, foi utilizado RNA proveniente de um painel de 20 tecidos normais (*Ambion*) para a construção do perfil de expressão.

4.2. Extração de RNA ^(Figuras 3 A.2 e 3B.2)

O RNA total foi extraído das linhagens através do método *TRIZOL Reagent* (*Life technologies, USA – Molecular Research Center, Inc*), segundo recomendações do fabricante. Para verificação de sua qualidade, o RNA foi aplicado em gel de agarose 1,5% e também quantificado a 260 nm em espectrofotômetro (*NanoDrop® 1000 Thermo Fisher Scientific*).

4.3. Síntese de cDNA ^(Figuras 3 A.2a e 3B.2)

A síntese de cDNA do RNA derivado das linhagem e painel de tecidos normais foi realizada com o kit de transcrição reversa *High Capacity cDNA Reverse Transcription Kit* (*Applied Biosystems*) conforme especificações do fabricante.

4.4. Análise dos níveis de expressão através de qRT-PCR ^(Figuras 3 A.2a e 3 B.2a)

Para a análise dos níveis de expressão dos genes *MGC16121* e *CR596471*, foram utilizadas sondas provenientes da *Integrated DNA Technologies* ou *Applied Biosystems* juntamente com *Taqman® Universal PCR Master Mix* (*Applied Biosystems*). Para a análise dos níveis de expressão dos miRNAs hospedeiros do gene *MGC16121* (miR-424 e miR-503) e outros miRNAs nas vizinhanças dos genes em estudo (miR-450a, miR-450b-5p e miR-542-3p), foram empregadas sondas provenientes da *Applied Biosystems* junto com *Taqman® Universal PCR Master Mix*, (*Applied Biosystems*).

As reações foram realizadas em replicata com amostras de cDNA diluídas no *7500 Fast Real Time PCR System*, produzido pela *Applied Biosystems*. O seguinte ciclo foi utilizado: 95°C (10 minutos), seguidos de 40 ciclos de 95°C (15 segundos) e 60°C (1 minuto). Foram utilizados os valores relativos à média geométrica dos Cts para os genes endógenos *GAPDH* e *HPRT* para normalização dos valores de Ct dos genes do estudo e dos snoRNAs *RNU24* e *RNU48* para os miRNAs em cada amostra analisada.

4.4.1. Análise de expressão dos genes *MGC16121* e *CR596471* e miRNAs da vizinhança em amostras de placenta ^(Figuras 3 A.1)

Com o intuito de verificar se havia expressão diferencial entre amostras de RNA de placentas em função do tempo de gestação, foram utilizadas 14 amostras de RNA de placenta cuja coleta e utilização já haviam sido aprovadas pelo Comitê de Ética em Pesquisa do HCRP da FMRP-USP, segundo o processo de número 9411/2010. Utilizou-se 3 amostras provenientes de gestações de 37 semanas, 5 amostras provenientes de gestações de 38 semanas e 6 amostras derivadas de gestações de 40 semanas.

Tais amostras foram verificadas para expressão dos genes e miRNAs do estudo conforme o item 4.4..

4.5. RACE (*Rapid Amplification of cDNA Ends*)^(Figura 3 A2.b)

Com a intenção de se obter a estrutura completa dos transcritos do *MGC16121* e *CR596471* e ainda verificar se havia variantes de *splicing* e poliadenilação, foi realizada a técnica de RACE. Tal técnica permite a amplificação das extremidades 3' e 5' das regiões não traduzidas dos genes. Com esta finalidade, utilizamos o *GeneRacer™ Kit* (*Invitrogen*). Todos os procedimentos descritos a seguir foram realizados de acordo com recomendações do fabricante.

A fonte de mRNA para este procedimento foi escolhida levando-se em conta o nível de expressão dos genes em estudo em amostras que já haviam sido analisadas até então. No caso do *MGC16121*, foi utilizado RNA extraído de HEK293 (linhagem de rim embrionário) e para o gene *CR596471*, RNA derivado UM-SCC-14 (linhagem de carcinoma de cabeça e pescoço). Tais amostras foram previamente tratadas com fosfatase alcalina intestinal de bezerro (CIP) e pirofosfatase ácida de tabaco (TAP) e, em seguida, ligadas a um adaptador (*RNA-oligo*) em sua extremidade 5' através da RNA ligase T4. Tanto o tratamento prévio quanto a ligação deste adaptador foram somente realizadas para obtenção da extremidade 5' (Figura 4).

Figura 4. Esquema das primeiras etapas da técnica RACE para obtenção da extremidade 5'. Modificado do manual de instrução do *GeneRacer™ Kit* (*Invitrogen*).

Para criar a fita complementar ao mRNA, foi utilizado o *primer Oligo dT*, complementar a cauda poli-A do mRNA. Nestas reações foi utilizada a transcriptase reversa *SuperScript™ II RT* (Figura 5).

Figura 5. Esquema da etapa de obtenção da primeira fita de cDNA. Extraído do manual de instrução do *GeneRacer™ Kit* (Invitrogen).

Para obtenção da extremidade 5', a primeira fita do cDNA foi amplificada usando *primers* específicos dos genes em estudo (*Reverse GSP*) (*CR596471*: 5' GCT GAA TAA CGG ATT ACC CC 3' e *MGC16121*: 5' GGA GTA CAG CCC ACT GTT TT 3') e um *primer* homólogo ao adaptador (*RNA-oligo*) ligado à extremidade 5' (Figura 6). Já para se obter a extremidade 3', a primeira fita de cDNA foi amplificada utilizando *primers forward* específicos dos genes em estudo (*Forward GSP*) (*CR596471*: 5' GCT GAG CGA AGC TTT TTG CT 3' e *MGC16121*: 5' GCC AGC CAG CCT TCC TGA AA 3') e um *primer* homólogo ao adaptador (*Oligo dT*) ligado à região 3' (Figura 7).

Figura 6. Esquema da etapa de obtenção da extremidade 5'. Extraído do manual de instrução do *GeneRacer™ Kit* (Invitrogen).

Figura 7. Esquema da etapa de obtenção da extremidade 3'. Extraído do manual de instrução do *GeneRacer™ Kit* (Invitrogen).

Nas reações para amplificar ambas as extremidades, foram utilizados os reagentes conforme recomendações do fabricante. Foi usada a enzima Platinum® *Taq* DNA Polymerase (*Invitrogen*) e empregado o seguinte ciclo: 1 ciclo de 2 minutos a 94°C; 5 ciclos de 30 segundos a 94°C e 1 minuto a 72°C; 5 ciclos de 30 segundos a 94°C e 1 minuto a 70°C; 25 ciclos de 30 segundos a 94°C, 30 segundos a 60°C e 1 minuto a 72°C e 1 ciclo de 10 minutos a 72°C. Posteriormente, 1 µl dos produtos destas reações foi utilizado em reações de *Nested-PCR* com os *primers* internos fornecidos pelo fabricante e os mesmos *primers* específicos citados. O ciclo utilizado nessas reações foi igual ao das primeiras.

Em seguida, os produtos das reações de *Nested-PCR* foram purificados após eletroforese em gel de agarose 2% por meio do *Gene Jet™ Gel Extraction Kit* (*Fermentas Life Science*) e clonados. Para a clonagem, foi utilizado o vetor *PCR®4-TOPO®* (*Invitrogen*) conforme recomendações do fabricante, utilizando-se bactérias eletrocompetentes *One Shot® TOP10 Competent Cells* (*Invitrogen*). Estas foram eletroporadas em cubetas (*Fisher Scientific*) a 2500 V no *Gene Pulser II®* (*Bio-Rad*) e logo em seguida adicionadas a 500 µl de meio SOC (*Invitrogen*) ou meio LB (Luria Bertani: 1% de tripton, 1% de NaCl e 0,5% de extrato de levedura) e deixados por 45 minutos em estufa a 37°C com agitação de 200 rpm. Posteriormente, foi adicionado mais 500 µl de meio às bactérias e 400 µl da mistura foram plaqueadas em placas com meio LB sólido contendo ampicilina 100 µg/ml e X-Gal. As placas foram deixadas *overnight* em estufa a 37°C.

4.5.1 Análise dos transformantes

As colônias brancas, que haviam sido transformadas com os fragmentos de interesse, foram repicadas em meio LB contendo ampicilina 100 µg/ml em placas de 96 poços. Estas foram deixadas em estufa a 37°C de um dia para outro e 1 µl do próprio meio foi utilizado para reação de PCR com a *Taq DNA Polymerase* (*Invitrogen*) com o kit do mesmo fabricante. O seguinte ciclo foi empregado: 1 ciclo de 3 minutos a 95°C; 35 ciclos de 40 segundos a 95°C, 40 segundos a 60°C e 55 segundos a 72°C e 1 ciclo de 5 minutos a 72°C.

Os produtos da PCR foram analisados em gel de agarose 2% e visualizados em luz ultravioleta com o transiluminador MiniBIS Pro (*DNR-Bio-Imaging Systems*). As

amostras cujas bandas correspondiam ao tamanho esperado dos fragmentos foram triadas por meio de sequenciamento no *MegaBACE 1000 DNA Sequencing System* (Amersham Biosciences) utilizando o DYEnamic™ ET Dye Terminator Kit (GE Healthcare Life Sciences), conforme recomendações do fabricante.

4.6. Estudo de metilação ^(Figura 3B)

Para os experimentos envolvendo a comparação dos níveis de expressão e metilação após o tratamento com agente demetilante, algumas linhagens as quais não expressavam ou pouco expressavam os transcritos dos genes em estudo (HCC1954: carcinoma de mama, e HT-29 e DLD-1: adenocarcinoma colorretal) foram utilizadas, sendo mantidas conforme o item 4.1. Inicialmente, foram plaqueadas $1,28 \times 10^5$ células por poço de placas de 6 poços. Após 24 horas, o agente demetilante 5-aza-2-deoxicitidina, DMSO (veículo), ou nenhuma substância foi adicionada ao meio das células. A concentração utilizada de 5-aza-2-deoxicitidina foi de quantidade suficiente para 5 μ M. O tempo do experimento foi de 96 horas, sendo que a cada 24 horas os meios eram trocados e adicionadas novas alíquotas do agente demetilante ou DMSO, após as 24 horas iniciais para a adesão das células.

Posteriormente, para a verificação da viabilidade celular, as células dos três grupos (controles e tratado) foram coradas com azul de Tripán e contadas em contador automático (*Countess automated cell counter* produzido pela *Invitrogen*).

Deste experimento, as células foram utilizadas para extração de RNA conforme item 4.2. e para extração de DNA conforme o item 4.7..

4.7. Extração de DNA e PCR após o tratamento com o 5-aza-2-deoxicitidina ^(Figura 3 B.3)

O DNA genômico das linhagens foi extraído através do método que se utiliza *TRIZOL Reagent* (*Life technologies, USA – Molecular Reasearch Center, Inc*), segundo recomendações do fabricante. Após a extração, o DNA foi quantificado em comprimento de onda igual a 260 nm em espectrofotômetro.

4.8. Modificação do DNA por bissulfito de sódio ^(Figura 3 B.4)

O DNA obtido segundo o item anterior foi submetido ao tratamento por bissulfito de sódio, que tem como base a desaminação de citosinas não metiladas a uracila e a manutenção de citosinas metiladas na presença de NaOH e bissulfito de sódio (Frommer *et al.*, 1992). Para este procedimento, foi utilizado o *EpiTect Bisulfite Kit* (*Qiagen*) conforme instruções do fabricante.

4.9. Análise dos níveis de metilação por MS-HRM (*Methylation-Sensitive High Resolution Melting*) ^(Figura 3 B.4a)

A técnica MS-HRM permite a análise em porcentagem dos níveis de metilação entre amostras de DNA modificadas por bissulfito de sódio e se baseia nas diferenças de tempo de dissociação do DNA dupla fita para DNA fita simples entre amostras diferencialmente metiladas após reação de PCR (Wojdacz & Dobrovic, 2007).

No presente trabalho, as porcentagens relativas de citosinas metiladas para cada gene proveniente das amostras tratada e não tratada foram determinadas por meio do monitoramento do sinal do corante que fluoresce quando intercalado ao DNA dupla fita após sua amplificação por PCR (Wojdacz & Dobrovic, 2007). Foi utilizado DNA padrão totalmente metilado e não-metilado fornecido pela *EpiTect Control DNA* (*Qiagen*) diluídos em variadas concentrações (0%, 20%, 50%, 80% e 100% metilado) os quais foram empregados como referência durante o procedimento. Os *primers* para a PCR inicial do procedimento foram desenhados na região das ilhas CpG identificadas na região 5' dos genes em estudo por meio do Genome Browse – UCSC (Feb. 2009, CRCh37/hg19) (Figura 8).

Figura 8. Ilhas CpG nas regiões 5' dos genes *MGC16121* e *CR596471*. As sequências correspondentes aos genes em estudo e suas respectivas ilhas CpG em suas regiões 5' estão representadas dentro dos retângulos em cor vermelha. As duas sequências do gene *MGC16121* representam suas isoformas e a sequência *LOC100506757* representa o gene *CR596471*. Modificado de Genome Browse – UCSC (versão: Feb. 2009, CRCh37/hg19).

O programa *Methyl Primer Express v1.0* (Applied Biosystems) foi utilizado para simular a conversão das sequências das ilhas CpG por bissulfito de sódio (citosinas não metiladas em uracilas e manutenção de citosinas metiladas em citosina).

Os *primers* foram desenhados manualmente levando em consideração a necessidade da presença de um número limitado de dinucleotídeos CpG que estejam o mais longe possível de sua extremidade 3' (neste caso, foi incluída a presença de somente um dinucleotídeo CpG em cada *primer*). Tal estratégia tem o intuito de diminuir o viés intrínseco da PCR em favorecer a amplificação de sequências não metiladas e relativamente pobres em dinucleotídeos CpG (Wojdacz *et al.*, 2009).

Para a ilha CpG situada na região 5' próxima ao gene *MGC16121* foi utilizado o seguinte par de *primers*: MG F MS-HRM: 5'GTTTATGCGTTT TAGTTTAGTTAGG3' e MG R MS-HRM: 5'CGTATTCCTACCACCAAATACC3', sendo que a temperatura de *annealing* utilizada foi de 60°C. Devido à grande dimensão da ilha CpG situada na região 5' próxima ao gene *CR596471*, esta foi analisada por dois pares de *primers*: CR F1 MS-HRM: 5'GAGGGTCGAGAAGGGTATTGG3' e CR R1 MS-HRM: 5'ATCCGTCCCAAAAACCTCTAAATAC 3', mais próximos a região 5' da ilha, (temperatura de *annealing* utilizada de 62°C); e os *primers* CR F2 MS-HRM: 5'CGGGGTGGGGATTTTGG3' e CR R2 MS-HRM: 5'ACAACTACGACCTCCCTC3', mais próximos a região 3' da ilha, (temperatura de *annealing* utilizada de 60°C).

Para a realização do MS-HRM, foram utilizados 10 ng de amostra de DNA modificado conforme o item 6.8, *MeltDoctor™ HRM Master Mix* (Applied Biosystems) 1X e 0,3 µM de cada sequência de *primer* (*forward* e *reverse*) para um volume final de 20 µl. As reações foram realizadas em replicata no 7500 Fast Real Time PCR System, (Applied Biosystems). O seguinte ciclo foi utilizado: 95°C (10 minutos), seguidos de 40

ciclos de 95°C (15 segundos) e 1 minuto correspondente à temperatura de *annealing*; seguidos da fase de *melting* correspondente a 95°C (10 segundos), 60°C (1 minuto), 95°C (15 segundos) e 60°C (15 segundos).

4.10. Caracterização do padrão de metilação das regiões 5' dos genes *MGC16121* e *CR596471* pelo sequenciamento das ilhas CpG^(Figura 3 B.4b)

As ilhas CpG das regiões 5' dos genes *MGC16121* e *CR596471* foram sequenciadas com o intuito de verificar a densidade de metilação de seus dinucleotídeos CpGs mesmas nas linhagens previamente analisadas por MS-HRM. Somente foram sequenciadas aquelas as quais apresentaram aumento significativo ($p < 0,05$) de expressão dos genes em estudo após o tratamento com agente demetilante (linhagens analisadas: DLD-1 e HCC1954 para *MGC16121* e DLD-1 para *CR596471*).

Visando cobrir uma área maior das ilhas CpGs estudadas, as mesmas amostras de DNA utilizadas para o MS-HRM foram amplificadas utilizando *primers* que flanqueavam regiões mais extensas das ilhas, visto que a técnica de MS-HRM só permite a amplificação de pequenos fragmentos. Neste trabalho os amplicons obtidos no MS-HRM tinham em média 150 pb. Para o sequenciamento, os *primers* foram desenhados com ajuda do programa *on-line* MethPrimer (Li & Dahiya, 2002).

Para a ilha CpG situada na região 5' próxima ao gene *MGC16121* foi utilizado o seguinte par de *primers*: Bis MG F: 5'ATATTTTATTTGAGTAGGGAAAGGG 3' e Bis MG R: 5'AATACAACAACAATTCATATTTTAAAATAT 3', sendo que a temperatura de *annealing* utilizada foi de 60°C. Já a ilha CpG situada na região 5' próxima ao gene *CR596471* foi dividida em duas partes e analisada por dois pares de *primers*: Bis CR1 F: 5'AGGAGAGGAGTTGTAGGGAA 3' e Bis CR1 R: 5'AAACTATAAACTCACCTATCCC 3', mais próximos da região 5' da ilha, (temperatura de *annealing* utilizada de 61°C); e Bis CR2 F: 5'TATTATGAGAGAGAGAATAGAAAGG 3' e Bis CR2 R: 5'CACCTAATTTTTTTTCCCCCA 3', mais próximos à região 3' da ilha, (temperatura de *annealing* utilizada de 60°C).

Para as reações de PCR foi utilizada a enzima Platinum® *Taq* DNA Polymerase (*Invitrogen*) com o kit do mesmo fabricante, conforme recomendações deste. O seguinte ciclo foi utilizado: 95°C (2 minutos), seguidos de 35 ciclos de 95°C (1 minuto), 1

minuto correspondente à temperatura de *annealing*, e de 72°C (1 minuto); seguidos de um ciclo de 72°C (7 minutos). Quando os produtos da PCR não eram visíveis, uma segunda PCR era realizada a partir de 1 µl da 1ª PCR. Para esta segunda PCR, as temperaturas de *annealing* foram aumentadas em 4°C para cada par de *primers*, o tempo de *annealing* também foi aumentado de 1 minuto para 1 minuto e 30 segundos, e foi adicionado 0,5 µl de DMSO a cada 25 µl da mistura de reação.

Os produtos das PCRs foram sequenciados de forma direta, sendo que para cada grupo do tratamento com o agente demetilante foram sequenciadas três amostras equivalentes a cada experimento independente realizado. Tal método direto, apesar de não ideal, foi preferível em relação à produção de clones contendo os produtos das PCRs, pois o sequenciamento por meio do *MegaBACE 1000 DNA Sequencing System*® (Amersham Biosciences), adequado para um maior número de amostras, não foi satisfatório comparado ao *ABI 3130 Genetic Analyser*® (Applied Biosystems), conveniente para um número menor de amostras. Como o sequenciamento de vários clones seria inviável por meio deste último, foi escolhido o método de sequenciamento direto no sequenciador que fornecia melhor qualidade para as sequências geradas. Desta forma, o sequenciamento foi realizado no sequenciador automático *ABI 3130 Genetic Analyser*® (Applied Biosystems) usando o kit *BigDye*® Terminator v3.1 Cycle Sequencing, segundo recomendações do fabricante.

4.11. Análise da conservação e estrutura secundária dos transcritos

As sequências de nucleotídeos obtidas pela técnica de RACE foram analisadas pela ferramenta *The Basic Local Alignment Search Tool* (NCBI-BLAST) (Altschul *et al.*, 1990), procurando-se sequências de RNA semelhantes contidas no *NCBI Reference RNA Sequences* (RefSeq_rna) (Pruitt *et al.*, 2005).

Em seguida, as primeiras sequências de RNA não humanas mais semelhantes obtidas foram comparadas aos dos genes em estudo, sendo alinhados ao genoma humano por meio da ferramenta *BLAT* (Kent, 2002 e website: <http://genome.ucsc.edu/>).

Posteriormente, as sequências foram analisadas estruturalmente e de forma comparativa às demais sequências semelhantes. Esta última análise foi realizada por meio do RNAstructure (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) utilizando o

algoritmo TurboFold (Harmanci *et al.*, 2011), que calcula estruturas comuns baseado em conformações que apresentam os menores valores de energia livre.

4.12. Forma de análise dos resultados

Os dados do perfil de expressão obtidos pela técnica de qRT-PCR foram normalizados com os dados relativos aos genes ou snoRNAs de expressão constitutiva e analisados pela fórmula $2^{-\Delta C_t}$. Tais dados, tanto derivados da expressão do perfil de linhagens tumorais, como de tecidos normais foram analisados para a Correlação de Pearson no programa *GraphPad Prism 4*. Já os dados de expressão referentes ao tratamento com o agente demetilante, além de serem normalizados, foram comparados entre amostras tratadas e não tratadas pela fórmula $2^{-\Delta\Delta C_t}$ (Schmittgen & Livak, 2001), sendo que a análise estatística (Teste de *Mann-Whitney*) foi realizada no programa *GraphPad Prism 4*.

Os dados referentes à comparação de expressão em amostras de placenta de diferentes tempos de gestação foram também normalizados e analisados pela fórmula $2^{-\Delta C_t}$. Neste caso, foi aplicada a análise estatística One-way ANOVA e pós-teste de Tukey realizadas também por meio do programa *GraphPad Prism 4*.

As sequências obtidas através da técnica de RACE e de PCR referente às ilhas CpG de amostras de DNA modificadas segundo o item 4.8 foram lidas com o programa *Codoncode Aligner* (CodonCode Corp., Dedham, MA) e alinhadas ao genoma humano por meio da ferramenta *BLAT* (Kent, 2002 e website: <http://genome.ucsc.edu/>). Já as sequências das amostras modificadas por bissulfito de sódio foram analisadas através do programa *on-line QUantification tool for Methylation Analysis (QUMA)* (Kumaki *et al.*, 2008). Tais sequências só foram analisadas quando a percentagem de conversão esperada alcançava o mínimo de 95%.

Em relação à técnica de MS-HRM, os resultados obtidos foram analisados através do programa *High Resolution Melt (HRM) Software v2.0* (Applied Biosystems).

5. RESULTADOS

5.1. Perfil de expressão dos genes *MGC16121* e *CR596471*, e miRNAs da vizinhança em um painel de tecidos normais

Um painel de vinte tecidos normais foi usado para avaliar o nível e o perfil de expressão dos genes *MGC16121* e *CR596471*. A expressão de *MGC16121* mostrou-se praticamente específica de placenta, sendo que a expressão residual em outros poucos tecidos foi considerada não relevante (Figura 9A). O gene *CR596471* igualmente foi mais expresso em placenta, entretanto foi encontrada expressão reduzida em células de ovário, coração, testículo e colo do útero; além de expressão residual em outros tecidos (Figura 9B).

Figura 9. qRT-PCR utilizando amostras de painel de tecidos normais. A. qRT-PCR para o locus *MGC16121*. B. qRT-PCR para o locus *CR596471*. Os genes endógenos *GAPDH* e *HPRT* foram utilizados para normalização das amostras.

Todos os miRNAs analisados (miR-424, miR-503, miR-450a, miR-450b-5p e miR-542-3p) foram mais expressos em tecido de placenta, porém foram observados níveis menores de expressão em geral em tecidos relacionados com a reprodução, tais como: ovário colo do útero, e testículo (Figura 10).

Figura 10. qRT-PCR utilizando amostras de painel de tecido normal para os miRNAs localizados próximos aos genes *MGC16121* e *CR596471*. Os snoRNAs *RNU24* e *RNU48* foram utilizados para normalização das amostras.

Para saber se os transcritos *MGC16121* e *CR596471* e os miRNAs miR-424 e 503, ambos hospedeiros do lócus *MGC16121*, são regulados pelo mesmo mecanismo, foi aplicada a análise de Correlação de Pearson pareando-se os valores de $2^{-\Delta Ct}$ dos transcritos em cada tecido (Figura 11). Os resultados revelaram correlação estatisticamente significativa entre todos os transcritos (Figura 11A-F).

Adicionalmente, foi feita a análise de Correlação de Pearson incluindo todos os outros miRNAs miR-450a, miR-450b-5p e miR-542-3p situados próximos do transcrito *MGC16121*. Do mesmo modo, a análise de correlação foi positiva todas as combinações (Apêndice).

Figura 11. Correlação de Pearson (r) entre diferentes loci em relação às expressões obtidas das amostras do painel de tecido normal.

5.2. Perfil de expressão dos genes *MGC16121* e *CR596471*, e miRNAs da vizinhança em linhagens tumorais

Nessa análise foram usadas 18 linhagens celulares de diferentes tumores (Figura 12). Considerando-se um valor de Ct limite de 34 para a detecção de expressão, o transcrito *MGC16121* estava expresso em 9 das 18 linhagens (50%) (Figura 12A), enquanto o transcrito *CR596471* foi expresso em todas as linhagens analisadas (100%) (Figura 12B).

Diferentemente dos tecidos normais, e em analogia com as linhagens tumorais, apenas o transcrito *MGC16121* foi mais expresso na linhagem tumoral de origem placentária JEG-3, enquanto que para o transcrito *CR596471*, ela foi a segunda menos expressa (Figura 12).

Figura 12. qRT-PCR utilizando amostras de linhagens tumorais. A. qRT-PCR para o lócus *MGC16121*. B. qRT-PCR para o lócus *CR596471*. Os genes endógenos *GAPDH* e *HPRT* foram utilizados para normalização das amostras.

O perfil de expressão dos miRNAs também foi heterogêneo entre as linhagens tumorais. A expressão destes foi maior ou estava entre as maiores na linhagem HEK293 e menor ou das menores para a linhagem HCC1954. Porém, o mesmo não foi observado para miR-542-3p (Figura 13).

Figura 13. qRT-PCR utilizando amostras de linhagens tumorais para os miRNAs localizados próximos aos genes *MGC16121* e *CR596471*. Os snoRNAs *RNU24* e *RNU48* foram utilizados para normalização das amostras.

Diferentemente do que ocorreu para o perfil de expressão nos tecidos normais, não houve nenhuma correlação significativa entre os dois miRNAs hospedeiros do transcrito *MGC16121* (Figura 14A), entre o miR-424 e *MGC16121* (Figura 14B) e entre o miR-503 e *MGC16121* (Figura 14C). Também não houve correlação entre o miR-424 e *CR596471* (Figura 14D); entre o miR-503 e *CR596471* (Figura 14E), e entre os dois transcritos dos genes em estudo (Figura 14F). Quando analisada a correlação incluindo os outros miRNAs da região, somente houve correlação positiva entre os miRNAs miR-503 e 450b-5p (Figura 15A), miR-503 e 450a (Figura 15B), miR-424 e 450a (Figura 15C), e miR450b-5p e 450a (Figura 15D).

Figura 14. Correlação de Pearson (r) entre diferentes loci em relação às expressões obtidas das amostras de linhagens tumorais.

Figura 15. Correlação de Pearson (r) entre combinações de miRNAs que foram positivas em relação às expressões obtidas das amostras de linhagens tumorais.

5.3. Comparação entre os níveis de expressão dos genes *MGC16121*, *CR596471* e miRNAs da vizinhança em tecidos normais e linhagens tumorais

Os dados relativos à expressão em tecidos normais e linhagens tumorais foram comparadas para avaliar se havia diferença estatisticamente significativa entre a média de expressão dos dois tipos de amostras. Conforme demonstrado na figura 16, os miRNAs miR-424, 450-a e 450b-5p apresentaram maior expressão em tecidos normais em comparação às linhagens tumorais ($P < 0,001$). Todavia, excetuando-se o miR-542-3p, todos os demais miRNAs e os genes *MGC16121* e *CR596471* apresentaram uma maior tendência de expressão em tecidos normais.

Figura 16. Comparação entre os níveis de expressão em tecidos normais e linhagens tumorais dos genes *MGC16121* e *CR596471*, e miRNAs da vizinhança. Cada ponto representa o valor de $2^{-\Delta C_t}$ para cada amostra utilizada nos itens 5.1 e 5.2. *** $P < 0,001$ (Teste de Mann-Whitney).

5.4. Análise de expressão dos genes *MGC16121* e *CR596471*, e miRNAs da vizinhança em amostras de placenta

Entre os genes *MGC16121* e *CR596471*, e miRNAs da vizinhança, somente o miR-542-3p apresentou diferença de expressão estatisticamente significativa entre a amostra de placenta da 37^a semana de gestação em relação as da 38^a e 40^a semanas (Figura 17). Com exceção do miR-424, todos os demais loci e miRNAs apresentaram apenas uma tendência na redução de expressão.

Figura 17. Comparação dos níveis de expressão dos genes *MGC16121* e *CR596471*, e miRNAs da vizinhança entre amostras de placentas obtidas em gestações de 37, 38 ou 40 semanas. *P<0,05 (One-way ANOVA).

5.5. Análise dos níveis de expressão após o tratamento com 5-aza-2-deoxicitidina

Após o tratamento com o agente demetilante, a viabilidade celular foi verificada. Pelo menos 74% das células analisadas estavam viáveis em todos os grupos de células considerados para este experimento.

A expressão do gene *MGC16121* foi significativamente aumentada (P<0,05) após o tratamento nas linhagens HCC1954 (1329,6 vezes) e DLD-1(38,5 vezes) (Figura 18).

Figura 18. Comparação dos níveis de expressão de *MGC16121* entre os grupos controle DMSO (veículo) e tratados com 5 μM de 5-aza-2-deoxicitidina (5-Aza-dC) em três linhagens. Os valores das médias foram obtidos de três experimentos independentes. Os genes endógenos *GAPDH* e *HPRT* foram utilizados para normalização das amostras. Foi utilizada a fórmula $2^{-\Delta\Delta CT}$, sendo considerada a amostra calibradora, o RNA derivado do grupo de células em que nada foi adicionado ao meio.* $P < 0,05$ (Teste de Mann-Whitney).

Os miRNAs miR-424 e 503, hospedeiros do transcrito *MGC16121*, também apresentaram expressão aumentada ($P < 0,05$) após o tratamento das linhagens HCC1954 (23,7 e 3,2 vezes, respectivamente) e DLD-1 (3,2 e 4,4 vezes, respectivamente) (Figuras 19 e 20).

Figura 19. Comparação dos níveis de expressão do miR-424 entre os grupos controle DMSO (veículo) e tratados com 5 μM de 5-aza-2-deoxicitidina (5-Aza-dC) em três linhagens. Os valores das médias foram obtidos de três experimentos independentes. Os snoRNAs *RNU24e* e *RNU48* foram utilizados para normalização das amostras. Foi utilizada a fórmula $2^{-\Delta\Delta CT}$, sendo considerada a amostra calibradora, o RNA derivado do grupo de células em que nada foi adicionado ao meio. * $P < 0,05$ (Teste de Mann-Whitney).

Figura 20. Comparação dos níveis de expressão do miR-503 entre os grupos controle DMSO (veículo) e tratados com 5 μ M de 5-aza-2-deoxicitidina (5-Aza-dC) em três linhagens. Os valores das médias foram obtidos de três experimentos independentes. Os snoRNAs *RNU24* e *RNU48* foram utilizados para normalização das amostras. Foi utilizada a fórmula $2^{-\Delta\Delta C_t}$, sendo considerada a amostra calibradora, o RNA derivado do grupo de células em que nada foi adicionado ao meio.* $P < 0,05$ (Teste de Mann-Whitney).

Houve aumento significativo da expressão do gene *CR596471* (4,1 vezes) (Figura 21) e do miR 450-a (4,4 vezes) (Figura 22) depois do tratamento com o agente demetilante na linhagem DLD-1.

Figura 21. Comparação dos níveis de expressão de *CR546971* entre os grupos controle DMSO (veículo) e tratados com 5 μM de 5-aza-2-deoxicitidina (5-Aza-dC) em três linhagens. Os valores das médias foram obtidos de três experimentos independentes. Os genes endógenos *GAPDH* e *HPRT* foram utilizados para normalização das amostras. Foi utilizada a fórmula $2^{-\Delta\Delta C_t}$, sendo considerada a amostra calibradora, o RNA derivado do grupo de células em que nada foi adicionado ao meio.*P<0,05 (Teste de Mann-Whitney).

Figura 22. Comparação dos níveis de expressão do miR-450a entre os grupos controle DMSO (veículo) e tratados com 5 μ M de 5-aza-2-deoxicitidina (5-Aza-dC) em três linhagens. Os valores das médias foram obtidos de três experimentos independentes. Os genes endógenos *GAPDH* e *HPRT* foram utilizados para normalização das amostras. Foi utilizada a fórmula $2^{-\Delta\Delta C_T}$, sendo considerada a amostra calibradora, o RNA derivado do grupo de células em que nada foi adicionado ao meio.* $P < 0,05$ (Teste de Mann-Whitney).

Por outro lado, não houve diferença significativa na expressão de todos os loci para a linhagem HT29 após o tratamento com o agente demetilante (Figuras 18 a 22). Os demais loci analisados (miR 450b-5p e miR 542-3p) também não apresentaram expressões alteradas após o tratamento em nenhuma das linhagens analisadas.

5.6. Análise dos níveis de metilação através de MS-HRM (*Methylation-Sensitive High Resolution Melting*)

Amostras de DNA convertidas com bissulfito de sódio (segundo item 4.8) derivadas das linhagens utilizadas no experimento envolvendo o agente demetilante 5-aza-2-deoxicitidina (conforme o item 4.6) foram analisadas para MS-HRM. Foram utilizadas amostras de DNA das linhagens cuja expressão dos genes em estudo aumentou significativamente após o tratamento com a droga demetilante.

Para a ilha CpG situada na região 5' do locus *MGC16121* (Figura 8), duas das 3 amostras de DNA correspondentes àquelas submetidas ao tratamento com 5-aza-2-deoxicitidina apresentaram-se 100% metiladas em ambas as linhagens analisadas (DLD-1 e HCC1954), padrão igual ao das amostras controle (Figura 23 e 24).

Figura 23. Padrão de metilação representativo de amostras correspondentes de dois dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem DLD-1 na ilha CpG a 5' do locus *MGC16121*. Controle 1 refere-se ao grupo em que nada foi adicionado ao meio e controle 2, em que somente o veículo DMSO foi adicionado. As setas, com suas respectivas cores, apontam as curvas correspondentes aos grupos do tratamento e curvas padrão.

Curvas de *Melting* normalizadas

Figura 24. Padrão de metilação representativo de amostras correspondentes de dois dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem HCC1954 na ilha CpG a 5' do locus *MGC16121*. Controle 1 refere-se ao grupo em que nada foi adicionado ao meio e controle 2, em que somente o veículo DMSO foi adicionado. As setas, com suas respectivas cores, apontam as curvas correspondentes aos grupos do tratamento e curvas padrão.

Todavia em uma das três amostras correspondentes aos experimentos independentes com o agente demetilante, as amostras de DNA cujas células foram submetidas à droga apresentaram-se menos metiladas em relação ao DNA das amostras controle. Tal resultado ocorreu para ambas as linhagens, sendo que para a DLD-1 a queda da metilação foi de 100 para cerca de 50%, e da HCC1954, de 100 para um valor entre 50 e 80% de metilação (Figura 25 e 26, respectivamente).

Curvas de *Melting* normalizadas

Figura 25. Padrão de metilação representativo de amostras correspondentes de um dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem DLD-1 na ilha CpG a 5' do locus *MGC16121*. Controle 1 refere-se ao grupo em que nada foi adicionado ao meio e controle 2, em que somente o veículo DMSO foi adicionado. As setas, com suas respectivas cores, apontam as curvas correspondentes aos grupos do tratamento e curvas padrão.

Curvas de *Melting* normalizadas

Figura 26. Padrão de metilação representativo de amostras correspondentes de um dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem HCC1954 na ilha CpG a 5' do locus *MGC16121*. Controle 1 refere-se ao grupo em que nada foi adicionado ao meio e controle 2, em que somente o veículo DMSO foi adicionado. As setas, com suas respectivas cores, apontam as curvas correspondentes aos grupos do tratamento e curvas padrão.

Para a ilha CpG situada na região 5' do locus *CR596471* (Figura 8), a região mais a 5' da ilha não apresentou mudança do padrão de metilação após tratamento com a droga demetilante para a linhagem DLD-1 em todas as amostras, sendo que estas apresentaram-se em torno de 0% metiladas (Figura 27).

Figura 27. Padrão de metilação representativo de amostras correspondentes de um dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem DLD-1 na região mais a 5' da ilha CpG a 5' do locus *C596471*. Controle 1 refere-se ao grupo em que nada foi adicionado ao meio e controle 2, em que somente o veículo DMSO foi adicionado. As setas, com suas respectivas cores, apontam as curvas correspondentes aos grupos do tratamento e curvas padrão.

O mesmo não foi observado para a região mais a 3' da ilha CpG mais próxima do gene *CR596471*. Houve diminuição da porcentagem de metilação em relação às amostras controle em uma das três réplicas experimentais da linhagem DLD1 após tratamento.. Neste caso, o padrão de metilação da amostra tratada foi de 20%, sendo que as do grupo controle foi de 100% (Figura 28). Como ocorreu para a ilha CpG próxima ao gene *MGC16121*, as outras amostras correspondentes ao grupo tratado não apresentaram variação no padrão de metilação em relação ao grupo controle (Figura 29).

Curvas de *Melting* normalizadas

Figura 28. Padrão de metilação representativo de amostras correspondentes de um dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem DLD-1 na região mais a 3' da ilha CpG a 5' do locus *C596471*. Controle 1 refere-se ao grupo em que nada foi adicionado ao meio e controle 2, em que somente o veículo DMSO foi adicionado. As setas, com suas respectivas cores, apontam as curvas correspondentes aos grupos do tratamento e curvas padrão.

Curvas de *Melting* normalizadas

Figura 29. Padrão de metilação representativo de amostras correspondentes de dois dos três experimentos independentes com o agente demetilante 5-aza-2-deoxicitidina (5-Aza-dC) para a linhagem DLD-1 na região mais a 3' da ilha CpG a 5' do locus *C596471*. Controle 1 refere-se ao grupo em que nada foi adicionado ao meio e controle 2, em que somente o veículo DMSO foi adicionado. As setas, com suas respectivas cores, apontam as curvas correspondentes aos grupos do tratamento e curvas padrão.

5.7. Caracterização do padrão de metilação dos promotores dos genes *MGC16121* e *CR596471* pelo sequenciamento das ilhas CpG

As figuras 30 e 31 mostram o resultado da média do padrão de metilação dentro das ilhas CpGs dos genes *MGC16121* e *CR596471* na linhagem DLD1 após tratamento com agente demetilante.

Para a ilha CpG situada na região 5' próxima ao gene *MGC16121* (Figura 30) foram verificados todos os 40 CpGs inclusos na região. Já a ilha CpG situada na região 5' próxima ao gene *CR596471* (Figura 31), a região foi dividida em duas partes, sendo que naquela situada mais longe do início de transcrição do gene *CR596471*, foram analisados 30 CpGs, e na outra parte da ilha, mais próxima a região 5' do gene, 35 CpGs. Desse modo, somente parte desta ilha pode ser verificada, já que seu tamanho total excedia a faixa na qual o sequenciador poderia identificar os dinucleotídeos. Além disso, a prévia análise dessas mesmas regiões por MS-HRM comprovou que somente uma parte dela poderia influenciar a transcrição do *CR596471* (item 5.6.). O uso de DMSO como controle não alterou o padrão de metilação, logo, o efeito deste veículo não foi considerado. Como esperado, houve a desmetilação das regiões analisadas das amostras tratadas em comparação as não tratadas.

Este fato não é evidente para a região 5' da ilha CpG situada no início do gene *CR596471*, como demonstrado também pelo MS-HRM, em que todos os grupos encontram-se próximos de 0% metilados (Figura 27). Talvez somente em sua região mais a 3', onde se nota que os dinucleotídeos CpG 28 e 29 estão menos metilados em média no DNA derivado das amostras tratadas do que nos grupos controle (Figura 31). A região 3' da ilha CpG próxima ao gene *CR596471*, aparentemente a região mais importante para a regulação deste locus, e a ilha CpG no início do gene *MGC16121* apresentaram-se totalmente metiladas para as linhagens sem prévio tratamento analisadas. Este fato é esperado, dado que tanto HCC1954 quanto DLD1 não expressam ou pouco expressam os transcritos dos genes em estudo, conforme o item 5.2..

Os dados da ilha CpG na região do gene *CR596471* gerados por MS-HRM e as médias dos níveis de metilação de cada dinucleotídeo CpG gerados por sequenciamento em relação ao grupo tratado foram diferentes (pelo MS-HRM um terço das amostras foram desmetiladas e pelo sequenciamento dois terços foram desmetiladas na região 3' da ilha). Isto se deve pelo fato de que uma das amostras de DNA do grupo tratado

analisada pelo MS-HRM foi posteriormente degradada. Sendo assim, uma amostra de DNA obtida de uma replicata foi utilizada para ser analisada por sequenciamento. Inesperadamente, esta última análise demonstrou que sua região mais 3' foi desmetilada, o que não havia sido observado para a amostra original analisada por MS-HRM.

Figura 30. Padrão de metilação de cada dinucleotídeo CpG da ilha CpG posicionada na regiões 5' do locus *MGC16121*. Os valores de metilação representam a média de três experimentos independentes. A. Representação gráfica da ilha CpG analisada. Modificado de Genome Browse – UCSC (Feb. 2009, CRCh37/hg19). B. Ilha CpG próxima ao *MGC16121* analisada para a linhagem DLD1; C. Ilha CpG próxima ao *MGC16121* analisada para a linhagem HCC1954. Controle - 1: grupo no qual nada foi adicionado ao meio; Controle - 2: grupo em que foi adicionado somente DMSO (veículo); 5μM 5-Aza-dC: grupo tratado com a droga demetilante 5-aza-2-deoxicitidina. ○ 0% metilado; ● 100% metilado; ◐ 75% metilado; ◑ 25% metilado.

Figura 31. Padrão de metilação de cada dinucleotídeo CpG da ilha CpG posicionada na região 5' do locus *CR596471* para a linhagem DLD1. Os valores de metilação representam a média de três experimentos independentes. A. Representação gráfica da ilha CpG analisada. Modificado de Genome Browse – UCSC (Feb. 2009, CRCh37/hg19). B. A área da ilha dentro do retângulo vermelho e quadro correspondente representam a região 5' da ilha CpG próxima ao *CR596471* analisada. C. A área da ilha dentro do retângulo verde e quadro correspondente representam a região 3' da ilha CpG próxima ao *CR596471* analisada. Controle - 1: grupo no qual nada foi adicionado ao meio; Controle - 2: grupo em que foi adicionado somente DMSO (veículo); 5μM 5-Aza-dC: grupo tratado com a droga demetilante 5-aza-2-deoxicitidina. ○ 0% metilado; ● 100% metilado; ◐ 75% metilado; ◑ 25% metilado.

5.8. Caracterização das extremidades 5' e 3' dos transcritos dos genes *CR596471* e *MGC16121*

Análise do transcrito CR596471

Para a caracterização da extremidade 5' e 3' do transcrito Utilizando-se RNA proveniente da linhagem UM-SCC-14, de carcinoma escamoso de cabeça e pescoço, foi realizada a técnica de RACE para o gene *CR596471*. Foram obtidas as extremidades 5' de duas isoformas diferentes do transcrito. Uma delas é constituída por 3 éxons (Figura 32) e outra, composta por 2 éxons (Figura 33). Para ambas as isoformas obteve-se 100% de identidade quando estas foram alinhadas ao genoma humano.

Figura 32. A. Sequência consenso derivada da técnica RACE para extremidade 5'. B. Representação gráfica do alinhamento com o *BLAT* (<http://genome.ucsc.edu>) (versão: Feb. 2009, CRCh37/hg19) da sequência consenso da região 5' da isoforma de *CR596471* que contém 3 éxons. Os nucleotídeos em vermelho representam os limites de cada éxon.

Figura 33. A. Sequência consenso derivada da técnica RACE para extremidade 5'. B. Figura correspondente ao alinhamento com o *BLAT* (<http://genome.ucsc.edu>) (versão: Feb. 2009, CRCh37/hg19) para isoforma de *CR596471* que contém 2 éxons. Os nucleotídeos em vermelho representam os limites de cada éxon.

A primeira tentativa de caracterizar a extremidade 3' do transcrito *CR596471* foi problemática, pois o primer intragênico (*Forward GSP CR596471*: 5' GCT GAG CGA AGC TTT TTG CT 3'), alinhava com uma região de elemento repetitivo localizada no éxon 1 do transcrito, gerando sequências com parte do intron 1 (Figura 34).

Figura 34. Elemento repetitivo obtido quando utilizado o *primer Forward GSP* complementar à região do éxon 1 do lócus *CR596471*. No interior do retângulo em vermelho é representada uma sequência consenso que aparentemente faz parte de um elemento repetitivo indicado no interior do retângulo verde. Figura correspondente ao alinhamento com o *BLAT* (<http://genome.ucsc.edu>, versão: Feb. 2009, CRCh37/hg19).

Visando contornar a influência deste elemento repetitivo na determinação da região 3' dos transcritos de *CR596471*, foi utilizado outro *primer Forward GSP CR596471* (5' GGG GTA ATC CGT TAT TCA GC 3') complementar à região do éxon 3', que por sua vez, se sobrepõe ao *primer Reverse GSP CR596471*: 5' GCT GAA TAA CGG ATT ACC CC 3', já utilizado para a determinação das extremidades 5' dos transcritos. Desta forma, foram obtidas duas sequências consenso mais abundantes referentes à extremidade 3' (Figura 35).

Figura 35. Sequências consenso obtidas a partir da técnica de RACE para a determinação das extremidades 3'. Foi utilizado um novo *primer Forward GSP* complementar à região do éxon 3 do lócus de *CR596471*. Figura correspondente ao alinhamento com o *BLAT* (<http://genome.ucsc.edu>, versão: Feb. 2009, CRCh37/hg19).

Alinhando-se as duas sequências consenso obtidas para a extremidade 3' com aquelas obtidas para a extremidade 5', foram geradas 4 sequências correspondentes a diferentes isoformas obtidas do gene *CR596471* (Figuras 36 a 39). Dessa forma, obteve-se duas isoformas contendo 3 éxons, sendo que uma delas possui a extremidade 3' mais longa e a outra, tal extremidade mais curta. O mesmo ocorre para a isoforma constituída por 2 éxons.

Figura 36. A. Sequência de 3 éxons com extremidade 3' mais curta de *CR596471*. A sequência consenso é derivada da sobreposição de sequências mais abundantes derivadas da técnica de RACE 5' e 3'. Os nucleotídeos em vermelho nesta sequência representam os limites de cada éxon. B. Figura correspondente à representação esquemática da estrutura éxon/intron da isoforma indicando a direção da transcrição pelos símbolos (+) no DNA, 5', 3' e +1, além da cauda poli-A em vermelho.

Figura 37. A. Sequência de 3 éxons com extremidade 3' mais longa de *CR596471*. A sequência consenso é derivada da sobreposição de sequências mais abundantes derivadas da técnica de RACE 5' e 3'. Os nucleotídeos em vermelho nesta sequência representam os limites de cada éxon. B. Figura correspondente à representação esquemática da estrutura éxon/intron da isoforma indicando a direção da transcrição pelos símbolos (+) no DNA, 5', 3' e +1, além da cauda poli-A em vermelho. Os nucleotídeos em preto são aqueles adicionais presentes na cauda 3' mais longa.

Figura 38. A. Sequência de 2 éxons com extremidade 3' mais curta de *CR596471*. A sequência consenso é derivada da sobreposição de sequências mais abundantes derivadas da técnica de RACE 5' e 3'. Os nucleotídeos em vermelho nesta sequência representam os limites de cada éxon. B. Figura correspondente à representação esquemática da estrutura éxon/intron da isoforma indicando a direção da transcrição pelos símbolos (+) no DNA, 5', 3' e +1, além da cauda poli-A em vermelho.

Figura 39. A. Sequência de 2 éxons com extremidade 3' mais longa de *CR596471*. A sequência consenso é derivada da sobreposição de sequências mais abundantes derivadas da técnica de RACE 5' e 3'. Os nucleotídeos em vermelho nesta sequência representam os limites de cada éxon. B. Figura correspondente à representação esquemática da estrutura éxon/intron da isoforma indicando a direção da transcrição pelos símbolos (+) no DNA, 5', 3' e +1, além da cauda poli-A em vermelho. Os nucleotídeos em preto são aqueles adicionais presentes na cauda 3' mais longa.

As isoformas obtidas não estão contidas no *The Basic Local Alignment Search Tool* (NCBI-BLAST) (Altschul *et al.*, 1990). No entanto, a isoforma constituída por 3 éxons obtida possui isoforma semelhante à sequência contida no *NCBI RNA reference sequences collection* (RefSeq) (Pruitt *et al.*, 2005) para o lócus *CR596471* (agora denominado *LOC100506757*), apesar de a posição de seu éxon 1 ser diferente (Figuras 32, 36 e 37). A isoforma que contém 2 éxons não possui nenhum registro de isoformas similares nestes mesmos bancos de dados.

Análise do transcrito MGC16121

A determinação da extremidade 3' do gene *MGC16121* foi realizada a partir de RNA derivado da linhagem HEK293, uma das linhagens que apresentou maior expressão na análise por qRT-PCR. A figura 40 representa a sequência obtida com maior número de clones. O *primer GSP* intragênico é complementar ao éxon 2 do transcrito.

Figura 40. A. Sequência consenso derivada da técnica RACE para extremidade 3'. B. Figura correspondente ao alinhamento com o *BLAT* (<http://genome.ucsc.edu>, versão: Feb. 2009, CRCh37/hg19), para o gene *MGC16121*. Os nucleotídeos em vermelho representam os limites de cada éxon.

Para a extremidade 5', pelo fato da dificuldade de isolamento da isoforma completa do gene na amostra de RNA da linhagem HEK293, foi determinada a partir de RNA da linhagem MCF-7 após tratamento com o agente demetilante 5-Aza-dC. Tal amostra já havia sido utilizada previamente para esta finalidade como projeto piloto e devido aos resultados mais satisfatórios obtidos com esta fonte de RNA, os clones produzidos foram reexpandidos a partir desta linhagem e utilizados para o sequenciamento conforme item 4.5.1. A figura 41 representa a sequência obtida mais abundante.

Figura 41. A. Sequência consenso derivada da técnica RACE para extremidade 5'. B. Figura correspondente ao alinhamento com o *BLAT* (<http://genome.ucsc.edu>, versão: Feb. 2009, CRCh37/hg19) para o gene *MGC16121*. Os nucleotídeos em vermelho representam os limites de cada éxon.

Obteve-se uma isoforma consenso de um dos transcritos do locus *MGC16121* a partir da sobreposição das sequências mais abundantes de ambas as extremidades determinadas (Figura 42). Tal isoforma possui uma extremidade 3' maior e uma extremidade 5' menor que aquela contida no *NCBI RNA reference sequences collection* (RefSeq) (Pruitt *et al.*, 2005).

Figura 42. A. Sequência consenso do transcrito de *MGC16121*. A sequência consenso é derivada da sobreposição de sequências mais abundantes derivadas da técnica de RACE 5' e 3'. Os nucleotídeos em vermelho nesta sequência representam os limites de cada éxon. B. Figura correspondente à representação esquemática da estrutura éxon/intron da isoforma indicando a direção da transcrição pelos símbolos (-) no DNA, 5', 3' e +1, além da cauda poli-A em vermelho.

5.9. Comparação dos perfis de expressão entre as isoformas do gene *CR596471* em um painel de tecidos normais e linhagens tumorais

Com o intuito de verificar as diferenças em relação ao perfil de expressão entre as duas isoformas encontradas, estas foram analisadas por qRT-PCR para o perfil de expressão em tecidos normais e linhagens tumorais conforme o item 4.4..

Como a isoforma constituída de 3 éxons, a nova isoforma encontrada (que contém 2 éxons) também foi mais expressa em placenta do que qualquer outro tecido normal do painel, porém menos expressa do que a isoforma de 3 éxons para todos os tecidos (Figura 43). O mesmo resultado ocorreu em linhagens tumorais, sendo a isoforma de 2 éxons também muito menos expressa do que a isoforma que contém 3 éxons em todas as linhagens analisadas (Figura 44).

CR596471: Isoforma de 2 éxons x Isoforma de 3 éxons

Figura 43. Comparação entre isoformas do *CR596471* em relação ao perfil de expressão em painel de tecidos normais por meio de qRT-PCR. Os genes endógenos *GAPDH* e *HPRT* foram utilizados para normalização das amostras.

CR596471: Isoforma de 2 éxons x Isoforma de 3 éxons

Figura 44. Comparação entre isoformas do *CR596471* em relação ao perfil de expressão em linhagens tumorais por meio de qRT-PCR. Os genes endógenos *GAPDH* e *HPRT* foram utilizados para normalização das amostras.

Apesar do menor nível de expressão da isoforma composta por 2 éxons, a análise de Correlação de Pearson entre ambas as isoformas para os níveis de expressão em amostras de tecidos normais e linhagens tumorais foi positivamente significativa (Figura 45).

Figura 45. Correlação de Pearson (r) entre as isoformas de *CR596471* em relação às expressões obtidas das amostras de tecidos normais (A) e linhagens tumorais (B). A análise foi realizada conforme o item 4.12.

5.10. Análise da conservação e estrutura secundária dos transcritos de *CR596471* e *MGC16121*

CR596471

O exame prévio para a identificação de sequências de DNA semelhantes no *NCBI Reference Genomic Sequences* (RefSeq_genomic) (Pruitt *et al.*, 2005) resultou no alinhamento com várias sequências de primatas, como o *Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*, *Pongo abelii*, *Nomascus leucogenys*, *Papio anubis*, *Macaca mulatta*, *Saimiri boliviensis* e *Callithrix jacchus* e em uma de ungulado, como o *Bos taurus*. Suas porcentagens de identidade com a região de DNA correspondente às isoformas de *CR596471* variaram conforme o grau de proximidade destas espécies com *Homo sapiens* (dados não mostrados).

A análise de ambas as isoformas de *CR596471* (isoformas contendo 2 ou 3 éxons) no banco de dados *NCBI Reference RNA Sequences* (RefSeq_rna) retornou sequências de RNA semelhantes encontradas em somente algumas das espécies citadas (Figuras 46 e 47).

Quando as sequências de *Pan paniscus*, *Pan troglodytes* e *Gorilla gorilla*, as mais similares, foram alinhadas com as isoformas aqui encontradas em humanos por meio da ferramenta *BLAT*, foi observado alinhamento semelhante para os éxons 2 e 3,

exceto para a isoforma de *CR596471* que contém 2 éxons (Figura 48). Os éxons número 1 dos transcritos humanos aqui obtidos possuem posições diferentes entre eles e também distintos dos transcritos dos outros primatas. Além disso, seus éxons terminais são menores que os dos transcritos semelhantes encontrados nos outros animais.

As figuras 49 a 53 mostram as estruturas secundárias conservadas para as cinco sequências. As estruturas dos transcritos *CR596471* revelaram-se mais semelhantes entre si do que aquelas homólogas das outras espécies. Estas, por sua vez, também são mais semelhantes entre si, assim como suas sequências.

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

	Description	Max score	Total score	Query cover	E value	Max ident	Accession
☐	Homo sapiens long intergenic non-protein coding RNA 629 (LINC00629), non-coding RNA	1011	1011	85%	0.0	100%	NR_038998.1
☐	PREDICTED: Pan paniscus uncharacterized LOC100983963 (LOC100983963), misc RNA	994	994	85%	0.0	99%	XR_156496.1
☐	PREDICTED: Pan troglodytes uncharacterized LOC100613576 (LOC100613576), misc RNA	994	994	85%	0.0	99%	XR_129508.1
☐	PREDICTED: Gorilla gorilla gorilla uncharacterized LOC101131505 (LOC101131505), misc RNA	983	983	85%	0.0	99%	XR_177265.1
☐	PREDICTED: Papio anubis uncharacterized LOC101023631 (LOC101023631), misc RNA	795	795	85%	0.0	93%	XR_163443.1
☐	PREDICTED: Pongo abelii uncharacterized LOC100451073 (LOC100451073), mRNA	172	172	14%	3e-40	100%	XM_002832138.1

Figura 46. Sequências de RNA semelhantes àquela encontrada para *CR595471* (sequência contendo 3 éxons) e seus detalhes. Figura correspondente à análise pelo *The Basic Local Alignment Search Tool* (NCBI-BLAST) (blast.ncbi.nlm.nih.gov/. Acesso em 12 de dezembro de 2012).

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

	Description	Max score	Total score	Query cover	E value	Max ident	Accession
☐	Homo sapiens long intergenic non-protein coding RNA 629 (LINC00629), non-coding RNA	793	873	72%	0.0	100%	NR_038998.1
☐	PREDICTED: Pan paniscus uncharacterized LOC100983963 (LOC100983963), misc RNA	782	857	72%	0.0	99%	XR_156496.1
☐	PREDICTED: Pan troglodytes uncharacterized LOC100613576 (LOC100613576), misc RNA	782	857	72%	0.0	99%	XR_129508.1
☐	PREDICTED: Gorilla gorilla gorilla uncharacterized LOC101131505 (LOC101131505), misc RNA	771	846	72%	0.0	99%	XR_177265.1
☐	PREDICTED: Papio anubis uncharacterized LOC101023631 (LOC101023631), misc RNA	634	634	62%	3e-179	95%	XR_163443.1
☐	PREDICTED: Pongo abelii uncharacterized LOC100451073 (LOC100451073), mRNA	403	403	34%	1e-109	99%	XM_002832138.1

Figura 47. Sequências de RNA semelhantes àquela encontrada para *CR595471* (sequência contendo 2 éxons) e seus detalhes. Figura correspondente à análise pelo *The Basic Local Alignment Search Tool* (NCBI-BLAST) (blast.ncbi.nlm.nih.gov/. Acesso em 12 de dezembro de 2012).

Figura 48. Sequências das isoformas de *CR596471* constituídas de 2 éxons (Homo2 na figura) e 3 éxons (Homo3 na figura) alinhadas pela ferramenta *BLAT* ao genoma humano (<http://genome.ucsc.edu>, versão: Feb. 2009, *CRCh37/hg19*) com as sequências de *Pan paniscus* (paniscus), *Pan troglodytes* (troglodytes) e *Gorilla gorilla* (Gorilla) obtidas no NCBI-BLAST (Figuras 46 e 47).

Figura 49. Estrutura secundária com menor energia livre do transcrito semelhante a *CR596471* obtido em *Pan paniscus*. A estrutura secundária foi obtida por meio do RNAstructure (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) utilizando o algoritmo TurboFold. Energia = 121,1.

Figura 50. Estrutura secundária com menor energia livre do transcrito semelhante a *CR596471* obtido em *Pan troglodytes*. A estrutura secundária foi obtida por meio do RNAstructure (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) utilizando o algoritmo TurboFold. Energia = 121.

Figura 51. Estrutura secundária com menor energia livre do transcrito semelhante a *CR596471* obtido em *Gorilla gorilla*. A estrutura secundária foi obtida por meio do RNAstructure (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) utilizando o algoritmo TurboFold. Energia = 121,9.

Figura 52. Estrutura secundária com do transcrito *CR596471* constituído de 3 éxons obtido em *Homo sapiens*. A estrutura secundária foi obtida por meio do RNAstructure (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) utilizando o algoritmo TurboFold. Energia = 59,1.

Figura 53. Estrutura secundária com menor energia livre do transcrito *CR596471* constituído de 2 éxons obtido em *Homo sapiens*. A estrutura secundária foi obtida por meio do RNAstructure (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) utilizando o algoritmo TurboFold. Energia = 59,5.

MGC16121

Assim como para a análise prévia feita para a conservação do DNA da região do *CR596471*, a mesma verificação foi realizada para o transcrito *MGC16121* no *NCBI Reference Genomic Sequences* (RefSeq_genomic). O resultado revelou identidade não só com os primatas (*Pan troglodytes*, *Pan paniscus*, *Gorilla gorilla*, *Pongo abelii*, *Nomascus leucogenys*, *Papio anubis*, *Macaca mulatta*, *Saimiri boliviensis* e *Callithrix jacchus* e *Otolemur garnettii*) e ungulados (*Bos Taurus*, *Ovis aries*, *Equus caballus* e *Sus scrofa*), mas também com felídeos (*Felis catus*), lagomorfos (*Oryctolagus cuniculus*), canídeos (*Canis lupus*), roedores (*Cavia porcellus*, *Rattus norvegicus*, *Mus musculus* e *Cricetulus griseus*) e proboscídeos (*Loxodonta africana*) (dados não mostrados).

A análise no banco de dados de RNA *NCBI Reference RNA Sequences* (RefSeq_rna) retornou seqüências de RNA semelhantes a *MGC16121* encontradas em

apenas algumas das espécies supracitadas (Figuras 54). Neste caso as sequências com mais identidade obtidas do banco foram das espécies *Nomascus leucogenys*, *Papio anubis*, *Saimiri boliviensis* e *Callithrix jacchus*. Quando foram alinhadas no *BLAT* ao genoma humano (Figura 55), foi observado que os últimos éxons foram aparentemente os mais conservados, sendo que em *Papio* somente o último permanece.

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

	Description	Max score	Total score	Query cover	E value	Max ident	Accession
☐	PREDICTED: Nomascus leucogenys hypothetical protein LOC100604482 (LOC100604482). mRNA	1229	1229	100%	0.0	98%	XM_003272604.1
☐	PREDICTED: Callithrix jacchus uncharacterized LOC100894936 (LOC100894936). mRNA	1120	1120	100%	0.0	95%	XM_003735825.1
☐	Homo sapiens uncharacterized protein MGC16121 (MGC16121). non-coding RNA	1029	1243	94%	0.0	100%	NR_024607.1
☐	PREDICTED: Papio anubis uncharacterized LOC101023398 (LOC101023398). misc RNA	976	976	82%	0.0	97%	XR_163442.1
☐	PREDICTED: Saimiri boliviensis boliviensis uncharacterized LOC101046212 (LOC101046212). misc RNA	931	931	84%	0.0	95%	XR_166171.1
☐	PREDICTED: Oryzomys latipes uncharacterized LOC100951067 (LOC100951067). misc RNA	706	706	84%	0.0	88%	XR_154290.1
☐	PREDICTED: Bos taurus uncharacterized LOC100847497 (LOC100847497). mRNA	691	691	83%	0.0	88%	XM_003588086.1
☐	PREDICTED: Bos taurus uncharacterized LOC100850934 (LOC100850934). mRNA	691	691	83%	0.0	88%	XM_003584676.1
☐	PREDICTED: Gorilla gorilla gorilla uncharacterized LOC101137390 (LOC101137390). mRNA	159	212	15%	3e-36	100%	XM_004064905.1
☐	PREDICTED: Pan paniscus uncharacterized LOC100990480 (LOC100990480). mRNA >reflXM_003954087.1 PREDICTED: Pan troglodytes uncharacterized LOC100990480 (LOC100990480). mRNA	159	212	15%	3e-36	100%	XM_003814521.1
☐	PREDICTED: Papio anubis uncharacterized LOC101023039 (LOC101023039). misc RNA	154	207	15%	1e-34	100%	XR_163441.1
☐	PREDICTED: Pongo abelii uncharacterized LOC100451073 (LOC100451073). mRNA	154	154	12%	1e-34	99%	XM_002832138.1

Figura 54. Sequências de RNA semelhantes àquela encontrada para *MGC16121* e seus detalhes. Figura correspondente à análise pelo *The Basic Local Alignment Search Tool* (NCBI-BLAST) (blast.ncbi.nlm.nih.gov/. Acesso em 12 de dezembro de 2012).

Figura 55. Sequência da isoforma obtida de *MGC16121* (Homo na figura) alinhada pela ferramenta *BLAT* ao genoma humano (<http://genome.ucsc.edu>, versão: Feb. 2009, CRCh37/hg19) com as sequências de *Nomascus leucogenys* (Nomascus), *Callithrix jacchus* (Callithrix), *Papio anubis* (Papio) e *Saimiri boliviensis* (Saimiri) obtidas no NCBI-BLAST (Figura 54).

As figuras 56 a 60 mostram as estruturas secundárias conservadas relativas às três sequências analisadas. Pôde ser observado que, com exceção daquele encontrado em *Papio anubis*, todos os outros transcritos apresentam pelo menos uma subestrutura semelhante (indicada na figura 56) localizada no último éxon. Sua sequência correspondente, retirada de *MGC16121*, é apresentada na figura 61.

Figura 56. Estrutura secundária com menor energia livre do transcrito de *MGC16121* obtido em *Homo sapiens*. A estrutura secundária foi obtida por meio do RNAstructure (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) utilizando o algoritmo TurboFold. A região no interior da elipse tracejada corresponde à subestrutura visualmente presente nas estruturas secundárias de todos os transcritos homólogos de *MGC16121* analisados (figuras abaixo), exceto o de *Papio anubis*. Energia = 64,8.

Figura 57. Estrutura secundária com menor energia livre do transcrito semelhante a *MGC16121* obtido em *Nomascus leucogenys*. A estrutura secundária foi obtida por meio do RNAstructure (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) utilizando o algoritmo TurboFold. Energia = 79,8.

Figura 58. Estrutura secundária com menor energia livre do transcrito semelhante a *MGC16121* obtido em *Callithrix jacchus*. A estrutura secundária foi obtida por meio do RNAstructure (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) utilizando o algoritmo TurboFold. Energia = 115,2.

Figura 59. Estrutura secundária com menor energia livre do transcrito semelhante a *MGC16121* obtido em *Saimiri boliviensis*. A estrutura secundária foi obtida por meio do RNAstructure (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) utilizando o algoritmo TurboFold. Energia = 79,2.

Figura 60. Estrutura secundária com menor energia livre do transcrito semelhante a *MGC16121* obtido em *Papio anubis*. A estrutura secundária foi obtida por meio do RNAstructure (<http://rna.urmc.rochester.edu/RNAstructureWeb/>) utilizando o algoritmo TurboFold. Energia = 80,0.

CATACCCACA	CTTCTTTGTT	CCAATTGGCA	TGTAAATTG	GTTTTCCGC	50
CAAATGAGTC	AGTCATGATG	GGAACCTCAA	CTGATTGAA	CAGATGTGTG	100
TCAATGTTAC	TTGGAAACT	AGATGTCAAT	AACCAGGGTC	ACAGAAAAAG	150
GCAGTGGTCA	CAACTCTGTA	AAAATGTATG	CATGCACACA	GACAAGAACT	200
AAAGTGAAC	CCCACACAGG	AAAACAGTGG	GCTGTACTCC	AGTGCTGGGA	250
CATTGAATGA	CTGTATG				

Figura 61. Parte da sequência do transcrito de *MGC16121*, correspondente à subestrutura secundária encontrada também nos transcritos de *Nomascus leucogenys*, *Callithrix jacchus* e *Saimiri boliviensis*.

6. DISCUSSÃO

6.1. Análise de expressão e correção

Um dos principais resultados revelados neste estudo foi a expressão elevada dos genes *MGC16121* e *CR596471* em placenta em comparação a outros tecidos normais. Ambos os genes são caracterizados como RNAs longos não-codificadores ou lncRNAs e estão localizados na região q26 do cromossomo X (Xq26), onde estão mapeados genes relacionados com a reprodução, sobretudo com o desenvolvimento fetal e placentário (Fant *et al.*, 2010; Jackman *et al.*, 2012). Os lncRNAs são também conhecidos como lincRNAs (large intergenic noncoding RNAs) e em sua grande maioria está expresso de forma tecido-específica, provavelmente regulando processos biológicos específicos em cada tecido (Sasaki *et al.*, 2007 e Cabili *et al.*, 2011).

Apesar da expressão aumentada dos genes *MGC16121* e *CR596471* na placenta, a grandeza do nível de expressão é baixa, como observada na figura 9A e B, onde os valores de $2^{-\Delta Ct}$ foram de aproximadamente 0.004 e 0.008, respectivamente. Para genes como o *PLAC1*, que está localizado na mesma região Xq26, os valores de $2^{-\Delta Ct}$ atinge níveis de 170 vezes maior na placenta (Silva *et al.*, 2007). Estes resultados corroboram com a evidência de que os lncRNAs geralmente são expressos em níveis menores que os genes codificadores (Cabili *et al.*, 2011). Visto que a composição celular da placenta é heterogênea, a expressão reduzida de alguns genes, como os lncRNAs, pode também ser explicada pela restrição da expressão a certos tipos de células (Sasaki *et al.*, 2007).

Diferentemente do gene *CR596471*, a expressão do gene *MGC16121* foi restrita à placenta, resultado que ainda não havia sido demonstrado na literatura. Segundo Sasaki *et al.*, (2007) e Cabili *et al.*, (2011), a maioria dos RNAs não-codificadores é expresso de forma tecido-específica, sugerindo papéis fisiológicos distintos em cada tecidos. No caso do gene *CR596471*, além da placenta, sua expressão foi também observada em órgãos relacionados à reprodução, como o cérvix (colo do útero), ovário e testículo. O mesmo padrão de expressão foi também observado quando analisamos a expressão dos miRNAs próximos aos genes estudados (miR-503 e miR-424, miR-542-3p, miR-450a e miR-450b-5p). Takada *et al.* (2006) já haviam demonstrado que em camundongos adultos, o mmu-miR-503 (mmj_157 no trabalho citado), homólogo do

miR-503 humano, é restrito à placenta e ovário. Tais resultados corroboram a sugestão de que a região Xq26 do cromossomo humano possa estar envolvida com a reprodução e desenvolvimento humano.

Segundo Baskerville & David (2005), miRNAs situados até 50 kb um do outro apresentam padrões de expressão altamente correlacionados, indicando que estes são processados de transcritos primários policistrônicos. A correlação positiva de expressões entre eles nos tecidos normais, descrita nesse estudo, indica que todos possam ser transcritos policistrônicos.

Em relação aos miRNAs intragênicos do *MGC16121*, (miR-503 e miR-424), os resultados de correlação entre eles e em tecidos normais foram positivas e sugere que ambos sejam derivados do transcrito *MGC16121* (Forrest *et al.*, 2010). Adicionalmente, a constatação de que a expressão do gene *CR596471* também é correlacionada com a do gene *MGC16121* e os miRNAs associados a ele, sugere que todos possam ser regulados por um mesmo agente controlador, pelo menos em condições normais.

Por outro lado, outra hipótese para a coexpressão de genes vizinhos poderia ser conferida à presença de uma região de cromatina aberta que envolvesse todo o local na qual eles se encontram (Batada *et al.*, 2007) e não a um transcrito policistrônico. Particularmente no caso de *MGC16121* e *CR596471*, dado que estão em orientações opostas e são relativamente próximos, a coexpressão poderia ser resultado da presença de um promotor divergente entre eles que, através da ligação de duas RNA polimerases, manteria a cromatina menos compactada (Seila *et al.*, 2009).

Com relação ao aspecto funcional, quanto maior o nível de coexpressão entre genes ligados, mais provável é sua importância (Batada *et al.*, 2007). Portanto, os altos valores de correlação encontrados neste estudo sugerem que a coexpressão dos transcritos analisados em diferentes tecidos poderia não ser somente consequência aleatória da transcrição pervasiva do genoma, mas sim ser importante do ponto de vista funcional. Aliás, os resultados indicam também que eles possam participar dos mesmos processos celulares (Batada *et al.*, 2007).

Interessantemente, o mesmo resultado não foi observado quando analisamos a correlação entre todos os transcritos (mRNAs e miRNAs) em linhagens tumorais. Nesse caso, podemos inferir que a ausência de correlação observada nas linhagens tumorais possa ser consequência das alterações genéticas e epigenéticas típicas das células tumorais durante a tumorigênese. O mesmo foi observado quando usamos amostras de

tecidos tumorais (dados não mostrados), excluindo assim possíveis interferências típicas de estudos *in vitro* com linhagens celulares.

Embora a diferença de expressão dos genes *MGC16121* e *CR596471* nos tecidos normais e linhagens tumorais não tenham sido estatisticamente diferentes, e o fato do *MGC16121*, nos tecidos normais, ser expresso especificamente em placenta e também em algumas linhagens tumorais destacou-se como um resultado relevante. Essas características podem caracteriza-lo como um bom biomarcador para doenças relacionadas a placenta, ou até como marcador de prognóstico de vários cânceres. Porém, devemos investir em estudos funcionais para elucidar essas questões.

Com relação aos miRNAs, nossos resultados revelam pela primeira vez uma maior expressão, de cerca de 10 vezes, dos miRNAs miR-424, miR-450a e miR-450b-5p ($p > 0,001$) nos tecidos normais em relação às linhagens tumorais (Figura 16). No caso do miR-450a, Castilla *et al.* (2011) revelaram uma expressão elevada em células mesenquimais quando comparado a células epiteliais de carcinosarcomas endometriais. Já a maior expressão do miR-424 em tecidos normais corrobora com os relatos do seu provável papel como supressor tumoral (Liu *et al.*, 2008; Pallasch *et al.*, 2009; Forrest *et al.*, 2010; Xu *et al.*, 2012;). Todavia, o papel dos miRNAs miR-450a e miR-450b-5p ainda é desconhecido. Os resultados são relevantes e também incluem esses miRNAs como candidatos a marcadores genéticos tumorais de diagnóstico, prognóstico e terapêutico, como já relatado para outros RNAs não-codificadores (Gibb *et al.*, 2011 e Pallasch *et al.*, 2009). Dados recentes, demonstram que os três miRNAs respondem a quimioterapia neoadjuvante (Svoboda *et al.*, 2012).

6.1.1. Análise de expressão das isoformas do gene *CR596471*

O estudo descreve pela primeira vez que o gene *CR596471* possui uma isoforma caracterizada pela ausência do segundo éxon. A maior expressão da isoforma contendo os três éxons, analisadas com sondas TaqMan específicas a cada uma delas sugere um papel de maior importância nos processos eventualmente regulados pelo gene *CR596471*. Como esperado, a diferença de expressão entre ambas deve ser regulada pela maquinaria de *splicing*, visto que a correlação de expressão tanto nas células normais como nas tumorais foram positivas (Figura 45).

6.1.2. Análise de expressão dos genes *MGC16121* e *CR596471*, e miRNAs da vizinhança em amostras de placenta

A tendência de expressão tanto dos genes *MGC16121* e *CR59647*, quanto dos miRNAs vizinhos analisados em placentas de diferentes tempos de gestação, sugere que todos possam estar envolvidos em processos biológicos que antecipam o fim do período gestacional (Figura 17). Uma análise focada nos genes alvos desses miRNAs podem apontar para os prováveis processos que estariam sendo regulados. Fant *et al* (2002) descreveram que o gene *PLAC1* possui níveis de expressão constantes entre a 22^a e 40^a semanas de gestação. Visto que este gene está localizado também na região Xq26, há a sugestão de que os mecanismos que regulam os genes desta região atuam de forma orquestrada durante a reprodução. Interessantemente, o miR-542-3p apresentou uma redução abrupta de expressão após 37^a semana de gestação. Tal miRNA deve ter uma atenção especial, pois a acentuada diferença deve revelar com mais clareza os processos biológicos que são regulados na placenta.

6.2. Regulação da expressão por metilação

A expressão seletiva de genes tecido-específicos pode ser regulada por meio da metilação do DNA (Esteller, 2003). Assim como para genes da família MAGE (Smet *et al.*, 1999; Sigalotti *et al.*, 2004), NY-ESO-1 (Oi *et al.*, 2009) e outros genes expressos especificamente em tecidos germinativos e em células tumorais, os genes *CR596471* e *MGC16121* também responderam a 5-aza-2-deoxicitidina em pelo menos uma linhagem tumoral analisada. Este resultado indica que tais loci são regulados, pelo menos em parte, pelo mecanismo de metilação do DNA.

O aumento significativo da expressão dos miRNAs hospedeiros de *MGC16121* após o uso do mesmo agente demetilante nas linhagens DLD-1 e HCC1954, corrobora a hipótese de que tais miRNAs e o gene *MGC16121* sejam corregulados. Adicionalmente, este resultado confirma o fato de que muitos miRNAs podem ser regulados de acordo com o grau de metilação de suas regiões promotoras (Lages *et al.*, 2012).

Através da análise de regiões das ilhas CpG na região 5' dos genes em estudo por meio de MS-HRM e sequenciamento foi possível verificar primeiramente que a região mais a 5' da ilha CpG ligada ao locus *C596471*, aparentemente não foi

responsável pelo aumento de sua transcrição. Além disso, tendo em vista que para todas as amostras analisadas nesta região o padrão de metilação estava em torno de 0% (Figuras 27 e 31), não haveria como a região ser responsiva ao agente demetilante.

Em contrapartida, para pelos menos um terço das amostras que foram submetidas ao tratamento com a 5-aza-2-deoxicitidina, houve diminuição da percentagem de metilação para a região mais a 3' da ilha CpG mais próxima ao locus do *CR596471* na linhagem DLD-1. Tal resultado também ocorreu para a região da ilha CpG a 5' do gene *MGC16121* nas linhagens HCC1954 e DLD-1. Duas entre três amostras tratadas com o agente demetilante mostraram-se 100% metiladas nas regiões analisadas e ainda assim mostraram-se hiperexpressas em relação às amostras controle.

Tais dados discrepantes poderiam ser explicados pela presença de DNA heterogêneo em relação aos níveis de metilação, mesmo após o tratamento com a droga demetilante, que por sua vez seriam derivados de populações heterogêneas de células tumorais, ou pela presença de 5-hidroxi-metilcitosina nas regiões das ilhas CpG analisadas. A conversão da 5-metilcitosina para 5-hidroxi-metilcitosina está associada à desmetilação (Tan & Shi, 2012) e este tipo de modificação não é discriminada pela alteração do DNA por bissulfito de sódio, utilizada nesse trabalho (Huang *et al.*, 2010).

6.3. Estrutura e conservação

As Etiquetas de Sequência Expressa (ESTs), como aquelas a partir das quais foram preditos os genes *MGC16121* e *CR596471*, são pequenas cópias dos transcritos de RNA que são muito propensas a conterem erros, principalmente em suas extremidades (Nagaraj *et al.*, 2007). Portanto, o método de RACE é o mais indicado para determinar a estrutura exata de um transcrito, incluindo a primeira e a última base a ser transcrita em determinado gene.

Todas as sequências dos transcritos gerados neste estudo não estão no banco de dados do *NCBI RNA reference sequences collection* (RefSeq) (Pruitt *et al.*, 2005), o que evidencia o fato de que grande parte dos lncRNAs apresentam muitas isoformas (Derrien *et al.*, 2012).

A determinação da extremidade 3' para os transcritos de *CR596471* foi dificultada pela presença de sítios de poliadenilação variantes contidos provavelmente em um elemento repetitivo (Figura 34). Tais elementos, além de poderem ser associados

à poliadenilação alternativa, podem modular a expressão tal como interromper a estrutura gênica (Parmigiani *et al.*, 2006).

A formação de mRNAs maduros em vertebrados depende da clivagem e poliadenilação do pré-mRNA, 10 a 30 pb após uma sequência sinal AAUAAA ou AUUAAA (Beaudoing *et al.*, 2000). Neste caso, os lncRNAs foram considerados como mRNAs no que tange ao término de sua transcrição, pois todos os transcritos aqui determinados possuem um sinal de poliadenilação e cauda poli-A. Os transcritos obtidos de *CR596471* apresentam o hexâmero de poliadenilação canônico AAUAAA (Anexos C e D), já o transcrito de *MGC16121* apresenta o sítio de poliadenilação menos comum, mas ainda prevalente, AUUAAA (Anexo B) (Ahmed *et al.*, 2011).

Outro elemento presente nos transcritos característico de genes codificadores são os sítios GT-AG, referentes aos sítios doadores-aceptores de *splicing* (Chodroff *et al.*, 2010) presentes nas limitações dos íntrons (Anexos B, C e D). Estes sítios canônicos estão presentes na maioria das junções intron/éxon, éxon/intron de lncRNAs (Derrien *et al.*, 2012).

Exceto para *MGC16121*, que apresenta homologia com uma sequência de RNA encontrada em *Bos taurus* (Figura 54), somente foram encontradas sequências homólogas aos transcritos dos genes estudados em primatas. Apesar deste fato, como citado no item 5.10, foi verificado que a sequência de DNA da região onde se localiza o gene *MGC16121* é conservada em vários outros mamíferos de diferentes grupos. Portanto, podem ser levantadas duas hipóteses no caso do gene *MGC16121*: este pode ter surgido como transcrito de forma independente em primatas e em *Bos taurus* ou simplesmente haveria falta de dados sobre este transcrito nas outras espécies citadas e, dessa forma, sua conservação seria subestimada. Adicionalmente, tanto a estrutura primária quanto a secundária dos transcritos analisados (*MGC16121* em *Homo* e sequências semelhantes em *Nomascus*, *Papio*, *Callithrix* e *Saimiri*) revelaram que o último éxon foi aparentemente a região mais conservada e, portanto provavelmente é a região mais importante e funcional dos transcritos.

No caso de *CR596471*, a ausência de sequências homólogas encontradas em outros mamíferos poderia sugerir que este gene tenha surgido em primatas, assim como muitos outros lncRNAs (Derrien *et al.*, 2012).

Apesar da não conservação total das sequências aqui encontradas com aquelas semelhantes em outras espécies, a estrutura secundária dos lncRNAs é mais conservada que a sequência de nucleotídeos (Liu *et al.*, 2012). Corroborando este fato, a análise das

sequências secundárias conservadas dos transcritos revela que há subestruturas visualmente parecidas, principalmente entre o transcrito de *MGC16121* e suas sequências semelhantes em *Nomascus*, *Callithrix* e *Saimiri*.

Para o *CR596471*, praticamente não há regiões conservadas entre humanos e outros primatas. As estruturas secundárias obtidas dos transcritos para o gênero *Pan* são conservadas, em razão de suas sequências primárias serem 100% idênticas. Já entre as isoformas aqui determinadas, há grande semelhança estrutural, principalmente na região do último éxon.

7. CONCLUSÕES

- O gene *MGC16121* é específico de placenta e também expresso em 50% das 18 linhagens tumorais analisadas;
- *CR596471* e miRNAs das vizinhanças (miR-424, 503, 450a, 450b-5p e 542-3p) são mais expressos em placenta do que qualquer outro tecido normal analisado, sendo o primeiro expresso também em 100% das linhagens tumorais avaliadas;
- Houve correlação positiva e significativa entre todos os genes e miRNAs em relação à expressão em tecidos normais, porém o mesmo não foi observado para linhagens tumorais;
- Os miRNAs miR-424, 450a e 450b-5p são significativamente superexpressos em tecidos normais comparados à linhagens tumorais;
- Com exceção do miR-542-3p, todos os outros miRNAs e genes avaliados são expressos de forma constante entre 37 e 40 semanas de gestação na placenta;
- *CR596471*, *MGC16121* e os miRNAs miR-424, 503 e 450a podem ser regulados negativamente por metilação do DNA;
- *CR596471* e *MGC16121* possuem ilhas CpGs situadas próximas às suas regiões 5' nos quais seus dinucleotídeos CpG são desmetilados quando tais genes são tratados pela ação da droga demetilante 5-aza-2-deoxicitidina;
- Foram determinadas as extremidades 5' e 3' e estrutura secundária de novos transcritos dos genes *MGC16121* (um transcrito composto de 3 éxons) e *CR596471* (um transcrito composto de 3 éxons e outro composto de 2 éxons). Tais transcritos, apesar de serem lncRNAs apresentam características de mRNAs, como presença de cauda poli-A, sinal de poliadenilação e sítios doadores e aceptores de *splicing*;
- A isoforma de *CR596471* composta por 2 éxons é menos expressa que a isoforma que contém 3 éxons tanto em tecidos normais, quanto em linhagens tumorais e suas expressões são positivamente correlacionadas;
- Os transcritos aqui determinados são relativamente conservados quando comparados a sequências de RNA encontradas em outros mamíferos, principalmente em primatas,

- O transcrito de *MGC16121* possui subestruturas secundárias visivelmente semelhantes com aquelas dos transcritos homólogos encontrados em *Nomascus leucogenys* e *Callithrix jacchus* e *Saimiri saimiri*, situadas no último éxon;
- O gene *MGC16121* pode ser considerado um possível bom marcador para diagnóstico, prognóstico e talvez para terapias contra cânceres;
- Os miRNAs miR-450a e 450b-5p foram relatados pela primeira vez como superexpressos em tecidos normais comparado a linhagens tumorais e portanto também possivelmente poderiam ser utilizados na medicina como marcadores na prevenção e tratamento de cânceres;
- Mais experimentos devem ser realizados para verificar a função dos genes *MGC16212* e *CR5976471*, além de avaliar mais robustamente a capacidade do gene *MGC16121* e miRNAs miR-450a e 450b-5p serem utilizados como ferramentas na medicina contra o câncer.

8.REFERÊNCIAS BIBLIOGRÁFICAS

AHMED F, BENEDITO V A & ZHAO PX. (2011). Mining functional elements in messenger RNAs: overview, challenges, and perspectives. *Frontiers in plant science*, 2:84.

AKERS S N, ODUNSI K & KARP A R. (2010). Regulation of cancer germline antigen gene expression: implications for cancer immunotherapy. *Future Oncology*, 6(5):717-32.

ALTSCHUL S F, GISH W, MILLER W, MYERS E W & LIPMAN D J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215: 403-410.

BARTEL D P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, (116): 281-297.

BASKERVILLE S & BARTEL D P. (2005). Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. (2005) *RNA*, (11): 241-247.

BATADA N N, URRUTIA A O & HURST L D. (2007). Chromatin remodelling is a major source of coexpression of linked genes in yeast. *Trends in genetics*, 23(10):480-484.

BEAUDOING E, FREIER S, WYATT J R, CLAVERIE J M & GAUTHERET D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome research*, 10(7):1001-1010.

BESTOR T H. (1998). The host defence function of genomic methylation patterns. *Novartis Found Symp.*, (214):187-95; discussion 195-9, 228-32.

BIRD A P. (1986). CpG-rich islands and the function of DNA methylation. *Nature*, (321): 209-213.

BUCK-KOEHN TOP B A & DEFOSSEZ P A. (2013). On how mammalian transcription factors recognize methylated DNA. *Epigenetics*, 8(2).

CABALLERO O L & CHEN Y T. (2009). Cancer/testis (CT) antigens: potential targets for immunotherapy. *Cancer Science*, (11):2014-2021.

CABILI M N, TRAPNELL C, GOFF L, KOZIOL M, TAZON-VEGA B, REGEV A & RINN J L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & development*, 25(18):1915-1927.

CASTILLA MÁ, MORENO-BUENO G, ROMERO-PÉREZ L, VAN DE VIJVER K, BISCUOLA M, LÓPEZ-GARCÍA MÁ, PRAT J, MATÍAS-GUIU X, CANO A, OLIVA E & PALACIOS J. (2011). Micro-RNA signature of the epithelial-mesenchymal transition in endometrial carcinosarcoma. *The journal of pathology*, 223(1):72-80.

CHODROFF R A, GOODSTADT L, SIREY T M, OLIVER P L, DAVIES K E, GREEN E D, MOLNÁR Z & PONTING C P. (2010). Long noncoding RNA genes: conservation of sequence and brain expression among diverse amniotes. *Genome biology*, 11(7):R72.

COCCHIA M, HUBER R, PANTANO S, CHEN EY, MA P, FORABOSCO A, KO M S & SCHLESSINGER D. (2000). *PLAC1*, an Xq26 gene with placenta-specific expression. *Genomics*, (68): 305-312.

COSTA F F. (2010). Non-coding RNAs: Meet thy masters. *Bioessays*, (7):599-608.

CROCE C M. (2009). Causes and consequences of microRNA dysregulation in cancer. *Nature Reviews. Genetics*, (10):704-714.

DANIEL F I, CHERUBINI K, YURGEL L S, DE FIGUEIREDO M A & SALUM F G. (2010). The role of epigenetic transcription repression and methyltransferases in cancer. *Cancer*, 117(4):677-687.

DERRIEN T, JOHNSON R, BUSSOTTI G, TANZER A, DJEBALI S, TILGNER H, GUERNEC G, MARTIN D, MERKEL A, KNOWLES D G, LAGARDE J, VEERAVALLI L, RUAN X, RUAN Y, LASSMANN T, CARNINCI P, BROWN J B, LIPOVICH L, GONZALEZ J M, THOMAS M, DAVIS C A, SHIEKHATTAR R, GINGERAS T R, HUBBARD T J, NOTREDAME C, HARROW J & GUIGÓ R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome research*, 22(9):1775-1789.

ESTELLER M. (2003). Relevance of DNA methylation in the management of cancer. *Lancet Oncology*, (6): 351-358.

FANT M, WEISOLY D L, COCCHIA M, HUBER R, KHAN S, LUNT T & SCHLES-SINGER D. (2002). *PLAC1*, a trophoblast-specific gene, is expressed throughout pregnancy in the human placenta and modulated by keratinocyte growth factor. *Molecular Reproduction and Development*, (63):430-436.

FANT M, FARINA A, NAGARAJA R & SCHLESSINGER D. (2010). *PLAC1* (placenta specific 1): a novel, X-linked gene with roles in reproductive and cancer biology. *Prenatal Diagnosis*, (30): 497-502.

FARAZI T A, SPITZER J I, MOROZOV P & TUSCHL T. (2011). miRNAs in human cancer. *The Journal of Pathology*, (223):102-115.

FARINA A, RIZZO N, CONCU M, BANZOLA I, SEKIZAWA A, GROTTI S & CARINCI P. (2005). Lower maternal plac1 mrna in pregnancies complicated with vaginal bleeding (threatened abortion <20 weeks) and a surviving fetus. *Clinical Chemistry*, 224-227.

FEINBERG A P & TYCKO B. (2004). The history of cancer epigenetics. *Nature Reviews Cancer*, 4(2):143-153.

FERRETTI C, BRUNI L, DANGLES-MARIE V, PECKING A P & BELLET D. (2007). Molecular circuits shared by placental and cancer cells, and their implications in the proliferative, invasive and migratory capacities of trophoblasts. *Human Reproduction Update*, 13(2):121-141.

FORREST A R R, KANAMORI-KATAYAMA A M, TOMARU Y, LASSMAN N T *ET AL.* (2010). Induction of microRNAs, miR-155, miR-222, miR-424 and miR-503, promotes monocytic differentiation through combinatorial regulation. *Leukemia*, (24): 460-466.

FRATTA E, CORAL S, COVRE A, PARISI G, COLIZZI F, DANIELLI R, NICOLAY HJ, SIGALOTTI L & MAIO M. (2011). The biology of cancer testis antigens: putative function, regulation and therapeutic potential. *Molecular Oncology*, (2):164-182.

FROMMER M., McDONALD LE, MILLAR DS, COLLIS CM, WATT F, GRIGG GW, MOLLOY PL, & PAUL CL. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA*, 89(5): 1827-1831.

GIBB E A, BROWN C J, & LAM W L. (2011). The functional role of long non-coding RNA in human carcinomas. *Molecular Cancer*, 10:38.

GRAVES J A. (2010). Review: Sex Chromosome Evolution and the Expression of Sex-Specific Genes in the Placenta. *Placenta*, (24):S27-S32.

HARMANCI A O, SHARMA G & MATHEWS D H. (2011). TurboFold: Iterative probabilistic estimation of secondary structures for multiple RNA sequences. *BMC Bioinformatics*, 12:108.

HIROTSUNE S, YOSHIDA N, CHEN A, GARRETT L, SUGIYAMA F, TAKAHASHI S, YAGAMI K, WYNshaw-BORIS A & YOSHIKI A. (2003). An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature*, (6935):91-96.

HUANG Y, PASTOR W A, SHEN Y, TAHILIANI M, LIU D R & RAO A. (2010). The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One*, 5(1):e8888.

HUNTZINGER E & IZAURRALDE E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature Reviews Genetics*, (12): 99-110.

HURST L & RANDERSON J P. (1999). An exceptional chromosome. *Trends in Genetics*, 15(10): 383-385.

HUSSAIN M U L (2012). Micro-RNAs (miRNAs): genomic organisation, biogenesis and mode of action. *Cell and Tissue Research*, (349):405-413.

JACKMAN S M, KONG X & FANT M E. (2012). Plac1 (Placenta-specific 1) is essential for normal placental and embryonic development. *Molecular Reproduction & Development*, 1-9.

JUNGBLUTH A A, SILVA WA JR, IVERSEN K, FROSINA D, ZAIDI B, COPLAN K, EASTLAKE-WADE S K, CASTELLI S B, SPAGNOLI G C, OLD LJ & VOGEL M. (2007). Expression of cancer-testis (CT) antigens in placenta. *Cancer Immunity*, (24):7-15.

JURKOWSKA R Z, JURKOWSKI T P & JELTSCH A. (2011). Structure and function of mammalian DNA methyltransferases. *Chembiochem*, 12(2):206-222.

KENT W J. (2002). BLAT- The BLAST-like alignment tool. *Genome Research*, (4): 656-664.

KONDO Y & ISSA J P (2004). Epigenetic changes in colorectal cancer. *Cancer and Metastasis Reviews*, 23(1-2): 29-39.

KONDO Y. (2009). Epigenetic Cross-Talk between DNA methylation and histone modifications in human cancers. *Yonsei Medical Journal*, 50(4): 455-463.

KOSLOWSKI M, SAHIN U, MITNACHT-KRAUS R, SEITZ G, HUBER C & TÜRECI Ö. (2007). A placenta-specific gene ectopically activated in many human cancers is essentially involved in malignant cell processes. *Cancer Research*, 9528-9534.

KUMAKI Y, ODA M & OKANO M. (2008). QUMA: quantification tool for methylation analysis. *Nucleic Acids Research*, (36): W170-W175.

KUSHI A, EDAMURA K, NOGUCHI M, AKIYAMA K, NISHI Y & SASAI H. (1998). Generation of mutant mice with large chromosomal deletion by use of irradiated ES cell – analysis of large deletion around hprt locus of ES cell. *Mammalian Genome*, (9): 269-273.

LAGES E, IPAS H, GUTTIN A, NESR H, BERGER F & ISSARTEL J P. (2012). MicroRNAs: molecular features and role in cancer. *Frontiers in Bioscience*, (17):2508-2540.

LI L C & DAHIYA R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics*, (11): 1427-1431.

LIU Q, FU H, SUN F, ZHANG H, TIE Y, ZHU J, XING R, SUN Z & ZHENG X. (2008). miR-16 family induces cell cycle arrest by regulating multiple cell cycle genes. *Nucleic Acids Research*, (16): 5391-5404.

LIU X, LI D, ZHANG W, GUO M & ZHAN Q. (2012). Long non-coding RNA gadd7 interacts with TDP-43 and regulates Cdk6 mRNA decay. *The EMBO journal*, 31(23):4415-4427.

LOPEZ J, PERCHARDE M, COLEY H M, WEBB A & CROOK T. (2009). The context and potential of epigenetics in oncology. *British Journal of Cancer*, (100): 571-577.

MASSABBAL E, PARVEEN S, WEISOLY DL, NELSON DM, SMITH SD & FANT M. (2005). PLAC1 expression increases during trophoblast differentiation: evidence for regulatory interactions with the fibroblast growth factor-7 (FGF-7) axis. *Molecular Reproduction and Development*, (71): 299-304.

MAZZIO E A & SOLIMAN K F. (2012). Basic concepts of epigenetics: impact of environmental signals on gene expression. *Epigenetics*, 7(2):119-130.

MITRA S A, MITRA A P & TRICHE T J. (2012). A central role for long non-coding RNA in cancer. *Frontiers in Genetics*, 3:17.

MORAN V A, PERERA R J, KHALIL A M. (2012). Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Research*, (14):6391-6400.

MURO E M, MAH N & ANDRADE-NAVARRO M A. Functional evidence of post-transcriptional regulation by pseudogenes. (2011). *Biochimie*, (11):1916-1921.

- NAGARAJ S H, GASSER R B & RANGANATHAN S. (2007). A hitchhiker's guide to expressed sequence tag (EST) analysis. *Briefings in bioinformatics*, 8(1):6-21.
- NUSSBAUM RL, MCINNES RR, WILLARD HF, THOMPSON MWT. (2001). *Thompson & Thompson Genetics in Medicine*. W. B. Saunders Company, Philadelphia.
- OI S, NATSUME A, ITO M, KONDO Y, SHIMATO S, MAEDA Y, SAITO K & WAKABAYASHI T. (2009). Synergistic induction of NY-ESO-1 antigen expression by a novel histone deacetylase inhibitor, valproic acid, with 5-aza-2'-deoxycytidine in glioma cells. *Journal of Neuro-Oncology*, (1): 15-22.
- OLD L J. (2007). Cancer is a somatic cell pregnancy. *Cancer Immunity*, (7):19.
- PALLASCH C P, PATZ M, PARK Y J, HAGIST S, EGGLE D, CLAUS R *ET AL*. (2009). miRNA deregulation by epigenetic silencing disrupts suppression of the oncogene *PLAG1* in chronic lymphocytic leukemia. *Blood*, (114): 3255-3264.
- PARMIGIANI R B, BETTONI F, VIBRANOVSKI M D, LOPES M H, MARTINS W K, CUNHA I W, SOARES F A, SIMPSON A J G, SOUZA S J & CAMARGO A A. (2006). Characterization of a cancer/testis (CT) antigen gene family capable of eliciting humoral response in cancer patients. *Proceedings of the National Academy of Sciences of United States of America*, 103(48): 18066-18071.
- PARRY L & CLARKE A R. (2011). The roles of the methyl-cpg binding proteins in cancer. *Genes & Cancer*, (2):618-630.
- PINHEIRO I, DEJAGER L & LIBERT C. (2011). X-chromosome-located microRNAs in immunity: Might they explain male/female differences? *BioEssays*, (33):791-802.
- PONTING C P, BELGARD T G. (2010). Transcribed dark matter: meaning or myth? *Human Molecular Genetics*, 19(R2):R162-168.
- PRUITT K D, TATUSOVA T & MAGLOTT D R. (2005). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, (33): 501-504.
- RAWN S M & CROSS J C. (2008). The evolution, regulation and function of placenta-specific genes. *Annual Review of Cell and Developmental Biology*, (24):159-181.
- RICE W R. (1984) Sex-chromosomes and the evolution of sexual dimorphism. *Evolution* (38):735-742.
- ROBERTSON K D. (2001). DNA methylation, methyltransferases, and cancer. *Oncogene*, (20): 3139-3155.

SASAKI Y T, SANO M, IDEUE T, KIN T, ASAI K & HIROSE T. (2007). Identification and characterization of human non-coding RNAs with tissue-specific expression. *Biochemical and biophysical research communications*, 357(4):991-996.

SAXENA A & CARNINCI P. (2011). Long non-coding RNA modifies chromatin: epigenetic silencing by long non-coding RNAs. *BioEssays*, 33(11):830-839.

SCHMITTGEN T D & LIVAK K J (2001). Analysis of relative gene expression data using Real-Time Quantitative PCR and the $2^{-2\Delta\Delta Ct}$ method. *Methods*, (25): 402-408.

SEARLE A G, EDWARDS J H & HALL J G. (1994). Mouse homologues of human hereditary disease. *Journal of Medical Genetics*, 31(1): 1-19.

SEILA A C, CORE L J, LIS J T & SHARP P A. (2009). Divergent transcription: a new feature of active promoters. *Cell Cycle*, 8(16):2557-2564.

SIGALOTTI L, FRATTA E, CORAL S, TANZARELLA S, DANIELLI R, COLIZZI F, FONSATTI E, TRAVERSARI C, ALTOMONTE M & MAIO M. (2004). Intratumor heterogeneity of cancer/testis antigens expression in human cutaneous melanoma is methylation-regulated and functionally reverted by 5-aza-2'-deoxycytidine. *Cancer Research*, (64): 9167-9171.

SILVA Jr W A, GNJATIC S, RITTER E, CHUA R, COHEN T, HSU M, JUNGBLUTH A A, ALTORKI N K, CHEN Y T, OLD L J, SIMPSON A J G, CABALLERO O L S D. (2007). *PLAC1*, a trophoblast specific cell-surface protein, is expressed in a range of human tumors and elicits spontaneous antibody responses. *Cancer Immunity*, (7):18-27.

SIMPSON A J G, CABALLERO O L, JUNGBLUTH A, CHEN Y T & OLD LJ. (2005). Cancer/testis antigens, gametogenesis and cancer. *Nature Reviews Cancer*, 5(8):615-25.

SMET C D, LURQUIN C, LETHE B, MARTELANGE V & BOON T. (1999). DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-Rich promoter. *Molecular and Cellular Biology*, (19):7327-7335.

SOUNDARARAJAN R & RAO A J. (2004). Trophoblast 'pseudo-tumorigenesis': Significance and contributory factors. *Reproductive Biology and Endocrinology*, (2):15.

SVOBODA M, SANA J, FABIAN P, KOCAKOVA I, GOMBOSOVA J, NEKVINDOVA J, RADOVA L, VYZULA R & SLABY O. (2012). MicroRNA

expression profile associated with response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer patients. *Radiation oncology*, 7:195.

TAKADA S, BEREZIKOV E, YAMASHITA Y, LAGOS-QUINTANA M, KLOOSTERMAN W P, ENOMOTO M, HATANAKA H, FUJIWARA S, WATANABE H, SODA M, CHOI Y L, PLAS-TERK R H, CUPPEN E & MANO H. (2006). Mouse microRNA profiles determined with a new and sensitive cloning method. *Nucleic Acids Research* 34, e115.

TAN L & SHI Y G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development*, 139(11):1895-1902.

The ENCODE Project Consortium. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, (447):799-816.

TOLLEFSBOL T. (2009). *Cancer epigenetics*. Boca Raton, London and New York.

TORARINSSON E, SAWERA M, HAVGAARD J H, FREDHOLM M & GORODKIN J. (2006) Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Research*, 16(7):885-889.

TSAI M C, SPITALE R C & CHANG H Y. (2011). Long intergenic noncoding RNAs: new links in cancer progression. *Cancer Research*, 71(1):3-7.

WAPINSKI O & CHANG H Y. (2011). Long noncoding RNAs and human disease. *Trends in Cell Biology*, 21(6):354-361.

WILSON A G. (2008). Epigenetic regulation of gene expression in the inflammatory response and relevance to common diseases. *Journal of Periodontology*, 79(8 Suppl):1514-1519.

WOJDACZ T K, BORGBO T & HANSEN L L. (2009). Primer design versus PCR bias in methylation independent PCR amplifications. *Epigenetics*, (4): 231-234.

WOJDACZ T K & DOBROVIC A. (2007). Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. *Nucleic Acids Research*, 35(6): e41.

XU J, LI Y, WANG F, WANG X, CHENG B, YE F, XIE X, ZHOU C & LU W. (2012). Suppressed miR-424 expression via upregulation of target gene Chk1 contributes to the progression of cervical cancer. *Oncogene*, (Apr 2): 1-12.

YAN B & WANGZ. (2012). Long noncoding RNA: its physiological and pathological roles. *DNA and Cell Biology*, 31(Suppl1):S34-S41.

9. APÊNDICE

Correlação de Pearson (r) para as 15 combinações restantes possíveis entre diferentes loci em relação às expressões obtidas das amostras do painel de tecido normal.

10. ANEXOS

ANEXO A

HOSPITAL DAS CLÍNICAS DA FACULDADE DE MEDICINA
DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO

Ofício nº 4319/2011
CEP/MGV

Ribeirão Preto, 23 de novembro de 2011

Prezados Senhores,

O trabalho intitulado "**CARACTERIZAÇÃO DA ESTRUTURA E DA EXPRESSÃO DOS GENES MGC16121 E CR596471**", foi analisado pelo Comitê de Ética em Pesquisa, em sua 337ª Reunião Ordinária realizada em 21/11/2011, e enquadrado na categoria: **APROVADO**, de acordo com o Processo HCRP nº 13867/2011.

Este Comitê segue integralmente a Conferência Internacional de Harmonização de Boas Práticas Clínicas (ICH-GCP), bem como a Resolução nº 196/96 CNS/MS.

Lembramos que devem ser apresentados a este CEP, o Relatório Parcial e o Relatório Final da pesquisa.

Atenciosamente.

DRª. MARCIA GUIMARÃES VILLANOVA
Coordenadora do Comitê de Ética em
Pesquisa do HCRP e da FMRP-USP

Ilustríssimos Senhores

BRUNA RODRIGUES MUYS

PROF. DR. WILSON ARAÚJO DA SILVA JÚNIOR

Depto. de Genética – FMRP-USP

ANEXO B

Sequência do transcrito *MGC16121* obtida incluindo os íntrons. As letras em maiúsculo representam os éxons e aquelas em minúsculo, os íntrons. Os sítios doadores-aceptores de *splicing* GT-AG estão representados em vermelho e sequência sinal de poliadenilação AUUAAA está indicada em verde. Retirado e modificado de *BLAT* (<http://genome.ucsc.edu>, versão: Feb. 2009, CRCh37/hg19).

```
GCTCCCCGCG AGGCCGGCTC AGGAGCAGgt aggtgggagg gggctcgcggt
ctctgtttcc ggcggggtcct cgcccttgct cacgcagcag ggctgtagtc
cggggagggc agtattcctg gctaggctgg ggcgcatagg cgggtgcgagt
cgaggagaga cgcggtgccc gcgctcagcc gtgccctagc agcgggaaca
gttctgcagt gagcgatcgg tgctctgggg tattgtttcc gctgccaggg
taagtctggg actgccgttc gggcctaccc agcaggtaag tcttaggccca
tcgcctgcgg tgtcggcccc gggagcagcg agaccaaga ttgcatcaaa
cgtttgcccc gtctctgcat atgtgactgg acgtgaatga tgcttcattt
tggcaccccc tttccctgct caggtagaat gtaaacaatt gaaaaaatc
ccttcccaat cagaatacca cccacaggac aacgaagaaa acctatcttc
agtccccacc cacgaaatgc tgggggggtgg ggaggactgg aaaaggagta
aggggagagg aaggggtgagc agctgcaaga agtctttaca cctctctgcy
gttcttggtta aacgcgctca tgtcctgggt atggaaggga cttagaattt
caggaaagat caaaataacg ggccctcttc cactccctc gccctccctt
ccccgctgct gggggggttag ggaggctccg tggagctcca gactctaga
aggtatcggt gaggcgcgag gccttgcygg cgggtgaacc caaggtgggg
tggaaggctc caaccgccc aatctgagcc cgaggcctgc tgagaagcat
gggggccttg ggatagtctc agaatgagga tgtgcgtttc tagctgcttt
gcgcctctct cccccaaaaa tctgtacca caattccaac ccggcggcac
gcccccaaga ctcttttgct gcccagggg cgggacctga gctgtcgatt
tcaggagccc ttcgtgactt caaaagtcct gggcactgtt gctcatgagt
gctgcacaa tgctgccttc taaagccacc tccatccctc actgggctgg
cctcctgagc cttcggtgag gaaacggggt tccgagttgc ccgcctgaga
gcttaacagt ctgactagaa aagggtctaat tcgctttctg tgcaaatctc
ttgagctaat tatttaatct gaaacatgga caggtaaagg accattggcg
ggcgtgggga gggagggtgt gaatcttttc cacacttttt gtcaccttaa
agaaaaaaaa aagacaccca aactcctctg gtctgaatgg gctgcctgga
actgggcttc ccggggccag gagtgggccc ctagaagttt ccagtttgg
ggtggggccg gggttccgaa gtctagagct gggagtcggg ttcagcggcc
cgagacgctg gctctcccag actcccttgc gccggggcag tgcgcggggg
cggtcgggt cctccccggg ccggcgggga cctcaggct ccgggggcgc
cgcccaccag cccgaccgg cggcctagaa atccccattc tgctcccgcc
tcgccccggc cggccagcgc agggccctc agggccgcaa tctccaggct
caaacggcga acaggtcccc gcatccgcct gcctttgcac ttgggggtgg
gagagccacc ttgtgacgcc cttctgcct ctcaccagct tcccccttc
tccccgtcct gtacggttcc ccttgagcc agAAGGAAGC CCGGTGCCAG
CCAGCCTTCC TGAAAGACCA AGCCCGCGCC ATCCGGCTTC CTCCAGTGA
CGCCTGCAGG ACCCAGgtga gacacggaac cgggttgggga cgcgagccag
ggatcctgac cgggttggggg cggcgagcaa gggatcctga cttgggggcy
ttgcctcccg gaatccaagc ggccttttgc ccgggtctcc cgagatgcat
```

Continuação ...:

```

cgaaattgag  ctcgccaca  cagccccctg  actgcgtgat  gcccttttct
gacttaaaac  attccttgta  cctttattta  cacgatccaa  gttgcagtgt
gcctagattt  tcctccaagg  acaaaaatgt  gaataaagca  atatttttagc
caaacatttc  tgtgtcaatt  aagaagattg  caattcgttc  cagactataa
attctcttta  acaactccct  aaggtattct  tttaaattgt  ttttcagcgt
tgctaaatgt  gcatcataat  gtgtgggaga  cttgtataaa  gtgaatcaat
ttataactcag  ctgattaatt  gacccgatgt  taccacctcc  aaggaggcac
ctttccacta  ttttctcatt  ttttcccaa  cacatgttat  ttgttagtga
ggaagctgcg  tgcacgtggc  attgtataaa  gtttattaga  cagacgtggc
ccttaaaacta  ggaagagctt  agtctaactc  taaatctaaa  attcgacatt
tcggtaagtt  gtgagacaag  ataccaggt  tgtttgataa  accagaattg
aacataattt  tttaaaaatt  taaatcacag  gaatgttttt  cttgaaggca
tccagcatct  ccagttagca  gtactgattt  tttttcccc  caacaaagGA
ACACTACATC  AACACTGTTG  GCGGGGACCT  GGACACAGAA  GACTCCTGTT
TCAAGAAAAT  ACAATCATCT  CTCAAAGGCT  GTAATTTATA  TGCATTTTAA
AACTCTAGGC  ATTGAAAACC  ACCCAAGTGT  CCCAAATAGA  AGGGTAATAT
ATAATCAATC  ACTCAGTGTA  ATATTATACA  TCCTTTAAAA  ATGTTATTGG
GAAATGTTTT  ATGATCTGTA  AGTCCAAGGA  ATCCTCTCCC  ACCATTTCTT
TCCCCCGCT  GTTCCCCCAT  ACCCACACTT  CTTTGTTCCA  ATTGGCATGT
AAATTTGGTT  TTCCCGCCAA  ATGAGTCAGT  CATGATGGGA  ACCTCAACTG
ATTTGAACAG  ATGTGTGTCA  ATGTTACTTG  GAAAACTAGA  TGTCAATAAC
CAGGGTCACA  GAAAAAGGCA  GTGGTCACAA  CTCTGTAAAA  ATGTATGCAT
GCACACAGAC  AAGAACTAAA  GTGGAACCCC  ACACAGGAAA  ACAGTGGGCT
GTACTCCAGT  GCTGGGACAT  TGAATGACTG  TATGCTGCTT  TTGATTTCCG
TTAATATGGG  CAACTGTCCA  ATTAAAAAAA  AAACCTCTAA  GGA

```

ANEXO C

Sequência do transcrito *CR596471* composto de 3 éxons e cauda poli-A mais curta obtida incluindo os íntrons. As letras em maiúsculo representam os éxons e aquelas em minúsculo, os íntrons. Os sítios doadores-aceptores de *splicing* GT-AG estão representados em vermelho e sequência sinal de poliadenilação AAUAAA está indicada em verde. Retirado e modificado de *BLAT* (<http://genome.ucsc.edu>, versão: Feb. 2009, CRCh37/hg19).

```

ACTGGGCGCC CAGAGTAACC AACACCGAGA CGAATAGCGA GGGCTCGGGG
AGAAAAAACG GAGGATGGCA AACGGCAAAG CCAAGGGCAC GAGtagcgc
agcgcgcacc cagaggctcc tgggacggac accaccccac gctcacgggt
cctctggcgg tgtcctttga gctgccggga cggcgcaagg tggctgagcg
aagctttttg ctcggtgctt acagggagta cctgagatgg tttgggaaag
gaaacagggg gtctcagacc gcttgggtgc agcgcctctc ccatgcgttt
gagtgcagat ctgtcgagac gccccagccc tcgccacccc accaagctct
caggtagaga gggcccgga cagggtgagc cccatagtcc tttggtgggc
ataacacaga gggaggagca attgtttaca catgtgcca gtacaagaag
ggaaactgat atttaagact gaagtgggga aacgggaggg aaaggagccc
ttatcctgga tgtagacgc tattaaggac catgctattt tgtatattgtg
agaatgatat tgaaggtaag gaggtcctta tccttgaaga tgcagtctga
agtgtacaga agtgaaatgt cagcatgtct gcagttttca aatgatgcag
ctaacatacg cacacattat gagagagaga acagaaagga agcaagtatt
gcaatatgtt aacagttgtt gcatcctgtt ggggggtggg ggtaaacaga
cgttcgttgt accgttcttt caacttgcc taaatatttgc aaaatgccat
aatgaaaatt ttggaagagt tcccttataa aatagtgcct gaatgctaag
aaaatgaacg cggcttcttt tcctttccct gtcccgggtg ggacaaatct
gagagcaagc agggctcaag gccagcgggg tctggacggg aagagcaatg
gtagagccg gcctgcagct gccatagccc tgggcaggcc aggaggggccc
atcgagccct cggggcagat aggtcgccgg gcgcaaaact atgcggagga
gcagccgagc gcgttggccc cgggggcggt gcgggggtgg gacttccctg
cggctctcag caacccttg tccgaccaga ctcggtgggc tcggctgcac
cacggtagag caaagaaacc ggcgggcccgc gggaaactgag ccgcgcgcct
gcagccgctc ccacggcggc ccgcggcctt ccccgcatc cccgagggga
ggcgtagatt gtcgcggcct gcagacagcg cgcagggtag cccagaaccg
cgtgcgggcg ctgggggaaa aaaaattagg tgccgaaaga aagcagcaac
gcacgcgcgc gcacacacac acgaaatcag tttacaaaaa taaaaccgcg
gaatctgccg tgctcggcgc ccgagtttga gctacttaac tctgcgatgc
atcttactgt catatctctg tagagctgat ggaaaatctg aaatgttttt
gcaagggcga tctcgagac ccaaggcggt gcaagcggtg cgcagccttt
cgctggcgct tcgggtcttt ttccggggcg cgaggaggga aaagagtagt
gtctttcagg gtctgcaaag ctgttttgaa aagttaacca aaaagggttt
aaatagcaag attttctttt caggaaaagt ttttaaacg gctgatactc
gtccagtttt ggtgggcgct aaaaagtctt attactaagc tacaagtttc
cctgttaaga ttttcaaaac ccctcgtgga cagggtgaa agtttgccgg
gctgtgagac gtgactgaac acgcttctaa atcctcagat agatcattta
taaagctctg gtgcccttg gatgaaaggc ttcgctgagt ttttaagact
tgcttcctg ggtaaatttg cttgggttgc tacacgtaag cattcatttc
cccacctaaa ttagtaaaag ggcagaagat tccaacgacg taagataaaa
gttgggttta tggcatgttc tctaaagaca aattgcaaca gctttcttta
ttaaacctct tttgctacct cagcaaattg gagagccagt tgaatattga
tgaattcttt gagggtgagt aatattgcta caggccggag aaagtaaggg

```

Continuação ...

```

aggacaagct taagcattta ctcaagtga caaccttcaa atgtttttca
actatgatgt ctcaaggatt ttcttaggac ctttgaagca aaccttgatt
ctgcaaattg cttcttattt ggtaggccct tcccaacctg acaaaactca
caggatataa attgccctaa gaagtgggta gaaacttacc ctagcattgg
aaacaggcac agcccaccac ctgggggtgt aggggggcct ccaactggta
gtcaggcaac ttctcacggg gcaactagga acctgaaatg aaaattgcac
accatctatt gttttttttg gtcccttctg aaatattgtt tcaaggcagc
tgttttcagg aaaatagcac atattttcct ttctcttttt gttccgtgaa
acgttcggaa atttatattga taaggaataa tcatagtccg ccgagggagc
aatagctttg aaggctctga aggcacagct gagaaaaata aattagctct
agactagaga tactcagcaa ggaagttagt cataacatgg tatttggcat
aaatttttga aaatgttttc tggaaagcaa ctgtgaatga atcttgggaa
atagatactg gagaatcaag ttgcattaaa acttattttt ttgtaaatta
caaagaagat ttctagaact ttccattttt attttaacta gaaatggtgt
ttaagattgt gtttttaatt aaggccttga gttagggaaa caagattttc
cttatgttgt aaatatccag aagtctttag agagataaaa tcaggataga
gttacaaatt taacagaaaag aaaacaactg gaattgtgtt atgaagcttc
tattgtgttt ttttgtaaag gctttttttc atggttcaag cagtggctag
cagtgtatag taactgtata ctataatata taatgtaaga atgcttgtcc
tactctcaga ttaaacaaaa aagacagtgt cctttggtaa tttttatttt
atgaaattag ttataaaaaat agttatttgc aaaactcaga actttgggag
aaaaactcta atcttaactt ctctacccc cgtgatccca tcaggaaaaa
gtagaatgaa ttagtaaat tttcctttct gccaaagaga agcttagtaa
gtcattagta gctgttgctt ttgatttcta gatctctttt ccctcctcaa
agacgcagaa tcaactgcctt cagtgtgccc tggatttctt ctttaaacca
ggcttttttt tttttttttc tgggtgtgta ctattattgt ctttaattatt
cctggatttt tgcccattaa atcaaaaatt gtctcagatg acggcaatga
ttggaacttt ttgctaagga gggcgagaat gaaggggtgt aggacagtct
cagcgtctct ttacaccgtg ctggccctat agtgcaacct gtaagtgggc
aggacttgct gtcctttgca cagagcgtga actgacagag caaaatgtaa
tgacaggagc atggagacca aaaacccatc tcttgtaggt gtggtggcca
taatggcggg tcaagtatag tactaggcta gatttagtca cttgctttca
ccataaaatt atggtaatca gtttcacctc acacttctag aactttcctc
gaacttcagc taattacgaa aaagagaaaa gaaaaacccc aatatatcag
tgtcaacctg tcaaattgcaa tcttgttgag ggaaactgaa gtgcctttgg
aaggcacttc attgatatca aaaaaattga aacaaacaaa caaaaacacc
aaccaaaagt gatcagcacc aaggatatcc tgtaaaaagc acagagctct
gctttgtgat gtctgaatga catcgctttt gggatcttgt tatgtgaaaa
gaaaacactc ctcggggcaa ataatatctt gctgggtaga aagaatagga
gggaacaaca aattgaagag tgctaaggga tgctatcctt gttcaaaaacc
aagatgcagc tggagtgttt tccaaagcat gaaatacatt gttgaaatag
cagagagaaa gtgaatgtgc atatctgaaa tgttcacatg agagatggta
acacggcatc atccagaact gaactcactc gtggttgcca ttctttcccc
tgaagctccg tttgcatttt aaaaaataaa acttaagaga tagagtgtca
attcgccggg catggtgggt gatgactgta atcccagcta ctcgggaggc
tgaggcagga gaattgcttg aaccggggag gtggagggtg tagtgaaccg
agatcgacc actgcactcc agcttggaac cgactctgtc aaacaaacaa
acaaacaaaa aaaacagcga tagagtgtct aggactgatg gataatgtca
tttctttttg ttctgtttt ttaaaattat tcttgggctt acgatcgggt
gggatttttg agttcatgac cagagtaagt cccagccctc cgtgtagaag
aggaagtaga aatccagatc ctgtttttta ttttagcatt tctcctaaac
ccagggtctg acagcttctt tagtccagag tccccaggaa tgctacgcc
ataggcaggg aagagtgaat atgagagtag ggtggagagc aagggccagg
gcagaagaca gaaagaaaga agatgaaaga agagaaaaaa aaaaacaagt
ccttctcctc tgtccctact atctgtcctc aagttcagag tctccacttg
ccacatgat gtcacaaagg tgaccttctc agggcccca ctgatctctc
agttgagttc tgctcctctc ccagtggtcc acggagttct gctgggaaaa

```

Continuação ...

cacaattctg	gtcatattaa	tcctcggcac	aaaaggcaga	tgactaaaaa
gcttaactcc	ttaatgtggc	attcagggca	ttctacaatc	tggctccaat
ctgcgcttct	gcttgcattc	cctaccctcc	agccacacct	tctttatgtg
tttatgctgg	ataccctgt	tcctcccatt	tccgcatcct	tcttaggcc
agctcaaagt	ccacctcttc	cttgaagcct	tctcagatgg	cctcagagtt
ggagttaact	gcctctgtcc	tgtcctgtca	cggcacttta	attaacttag
ttgttctgaa	attctcttcc	tagcagacag	taagctccta	gagggtaggg
acatgtctgt	ctataggtat	cttctcagcg	ctttgcacaa	taatttgcatt
ccgttgatgt	ccaatacata	tttgtttttt	tgttttgttt	tttttgtttg
ttttttgaga	tagaatctca	ctctgtcgcc	cagcctggag	tacagtgggtg
ccatctcggc	tcactacaaa	ctccgcctcc	caggttcaag	caattatcat
gcctcccaaa	tagctgtgat	tacaggcacg	caccatcaca	ctcggctaatt
ttttgtattt	ttagtagaaa	cggggtttca	ccatgttgac	caggctgggtc
tcaaactcct	gacctcaagt	gattcacctg	ccttggcctc	ccaaagtgtc
ggaattacag	gcatgagcca	ctgtgcccag	ccaagttaga	aagaaatcca
atacatattt	gttaactgaa	cgtatacctg	agaagggaaa	aaagggaaccg
gaagaagtga	gaacaagggtg	ctgtgctggc	ggtaaagatg	atggctaaca
tcctcaagag	ttggcttagg	caccattttg	tgttcttggg	gctgaccaag
ggttggttag	ctcatatcta	acatgttggc	ttcacaacaa	tccatctggg
actgacactt	ggaacacatc	ctctggaata	atatagaatt	aatttgcctt
taaattagct	gctcagtact	ataaactctc	tcctgaaaat	ctatcaccat
cattttatac	ttttaagctt	caagctgatc	cttgaaggga	aggtggaatt
tggatggaat	gggcaaaaaca	gcatttttcag	aggccaggag	gcagagaaga
actgaaagga	gatggacctt	gttggacaga	gccttttctac	tggggaatta
gagactgggc	tggacaagta	gggtgatgtc	agattatttg	gggggtgggga
ggcaggcgcg	gtggctcacg	cctgtaatcc	cagcactttg	ggaggccgag
acgggtggat	tgcttgaggc	caagagttca	agaccagcct	ggccaacatg
gtgaaaaccc	atctctacta	aaaaatacaa	aaactagccg	ggcatggttg
cgtgtgcctg	taatcccagc	tactcgggag	gctgaggcag	gagaatggag
agtagcttaa	acccaggggg	cagaggttgc	agtgaagtga	gatcgtgcca
ctgcactcca	gcctgggtga	ctgagactgt	ctcaaaaaaa	aaaaaaaaaa
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaagaggct	gggcatggtg
actcacgcct	gtaatcccag	cacttttggg	ggccaaggca	gggtggatcac
gaggtcagga	gttcaagacc	atcctgacca	agatgatgaa	accccgctctg
tactaaaaat	acaaaaatta	gccgagtatg	gtggcaggcg	cctgtaatcc
cagctactca	ggaggctgag	acaggagaac	tgtttgaacc	cgggtagcag
aggttgacgt	gagccgagat	tgcactcactg	cgctctagcc	tgggcgacag
actgagactc	tgtctcaaaa	aaaaaacaaa	aaacattatg	gggggtctgg
aagtaggcag	aggagtgtg	acttttagttg	caggatggta	tagaaccttt
tttttttttt	tttttttttt	ggtgtgatgc	ttttaaagg	gcaggggctg
gatgtaatga	aagccgtgtt	ttttaacaga	ttcacctggc	ttatatggta
taaccctatg	aactacaaac	tatggtaaac	ctaggactat	ttgaataatt
agaatatccc	attctccatg	ccacactgga	gtgaactttc	ttatttgaga
agataaagag	tggaaaaata	ctattttatca	tatacctatt	atgtatcagg
agatgtactg	gattcttccc	ccaccccagg	tattccctcc	ctctctcccc
cttccctccc	tcctctctct	acacacttgc	acttgcacac	accagaaatc
ccaactagga	cttgatttta	attgaccttt	aaagtaatca	attttttgtt
aaattactca	tacaaagcat	gctatgaaat	accagtaaat	atcgggttaa
aaattaactg	aaagacagaa	acttactatt	gtgtgctgat	ttgatggaac
agaggagtga	aaatatagta	cgcaccttct	ggattgagac	tgtggttttc
gtttttgttt	ttgttttttt	gtagtggcag	agcttcaaac	tggagaatgt
tctttttcat	gcgcagtttt	taaaaattgt	cattaaatat	tttcacaaag
ttaaaacaac	attgaaagta	acttgggcaa	cttagatatt	ccctgtactt
gctgtaactc	aaataccaaa	acataatttag	agtttaggtg	catctatcat
tcgaggcatg	tgattttctt	acctgggtga	agcgaaagaa	agttaaatcg
atgaaatatg	aaatgaaaat	gagaatcaat	ttgtagaatg	aaccataaaa
gaacgagatt	tttgaagaga	ttatttttatt	ttgattgcag	actgtaaat

Continuação ...

```

atcacagga gttgctgtag gaatgacatt gaggcttact tttgggtcaa
tttgttcctg tcctgaccca cccttggcag tggagacagc tagaatttct
ctaactccat agaaaccagt ctactgttg ctttaggact cgattctctc
cctacactgg gcttccttta gtgctttgtt atgtaaatac agcttcaaaa
atagaataaa tattctgtag gcattacttg gaaaacgagc ttttatttta
aagatgtgtg cttattttaag tcgcagacag agcttttaaag agctcttgat
agaaagaaat atatttggcg aagagtcaaa acttttaaag taactttaaa
aggtgaataa tttacattcg tatagtgttt acataatagg gtgtgtacga
caaagccaaa acaacattga aaataacctg agcaatttaa atattcttcc
tagtaactgt aaaataaaaag tgccttcgaa tggtagattt gctggcagga
aatttttttc atgttcccag gtggcaagga ttctttcatg ttcccttata
tttgaaaact ggtattagcg tttacttctc ggactcagac caattcccc
ttacctgtca gagaacatct agcaataaat gctgacaaag ctacatagca
aagggcacac ctgtcacttc cttttcagAA ACCTGTTTAT GCACCCTCCT
GCAATCCAC CACCTGCCCA GGCCAAGTGA GCCCTCTCC GCTACTGGTG
TTTCATAATG CTGGGAGCTA CTTTGATTTC ATGCATTGCC CTTCAgtaa
gccacaggaa cctcaccact ggccatgcc aatgggtgtg aattggataa
gaacaaagaa aaatggaagt agcaagtggg ggggagttaa gtgtactata
tgacttactc atatgcaagt caataagcaa agtcccctgc ctcttgtcct
tggtatgtgg ataataataa atataaatat acataatata tatgctgaca
tatatacatg catgcacaca cacacagaca cacattacct ggatccccc
cctcctcacc ccactttgaa aaactttaga gtatcggttg cctcatcaga
attacattac attagaggat tggtgaaaga tttcatatta ctgctaattg
ggggaacata attagaacca acgttagcgc ctctactcag tatcatcatt
gctattatth tattttacatc ttatgctttt ttcccttggc tttactgtc
ttcatcaatc tcctgttata aaaataaata aaagcttcc atggaattat
tttcaccagt caagtctgt aacttgtttt ccttgaagag ggaaaatata
actgatggac acaaggaaac agcagtggca tagcatgtcc tttccacccc
agtctgtgtt ttgacatcag ttgtatttac acacataaac tgaagaacgt
tgcatthgtg gagtttgcca gtgtggtaag attgaagggg tggtaggcag
cctctatgct ggtcccccact gatccctgcc ttctgggtgt cccactgttc
tgtaatccct cctagtacgt gcgctggacc tagagactct tctaacaagt
agaatatggc agaagccatg ggatgtcact tctgtggctt ctaacttggg
ccttttgtca tccccctccc cttctctcta tcagCTGCCA TGTTATGAGG
ACACTGGAG AAAGCTACAG GAGAGGAAGT GAGGTCTCCT GCCATCAGCC
ACTTGAGTGA GCATGGGAGT GGATTCTCCA AACTCCAGTT AAGACTTGAG
ATGTCCACAG TGGCAGCCAC TAGCTTGACT GCAACTTACG AGAGACCCTG
AGCCAGAGGC ACTCAGTTAA GTCATACACA GTTTCCTGAC CCGTAAAACC
TGTGACACAA TGAAGGTTTG TTGTTTTGGG GGTAATCCGT TATTGAGCAA
TAGATAACGA ATACAGAAGG CTTGTAATTG TATAACCAAC GTGAGTTTAT
AAGCGGATAT CTGACCTCAT TTGTTTTCTC CTGAAAAAGT TATAGAAAAA
TCACAAGACT GCAAGTCACT CTCTCTCTTT CTTTTGTGGG GAAATAACAA
GTATAATTAC AA

```

ANEXO D

Sequência do transcrito *CR596471* composto de 2 éxons e cauda poli-A mais curta obtida incluindo o íntron. As letras em maiúsculo representam os éxons e aquelas em minúsculo, o íntron. Os sítios doadores-aceptores de *splicing* GT-AG estão representados em vermelho e sequência sinal de poliadenilação AAUAAA está indicada em verde. Retirado e modificado de *BLAT* (<http://genome.ucsc.edu>, versão: Feb. 2009, CRCh37/hg19).

```

GAGAAGGGCA CTGGGCGCCC AGAGTAACCA ACACCGAGAC GAATAGCGAG
GGCTCGGGGA GAAAAAACGG AGGATGGCAA ACGGCAAAGC CAAGGGCACG
AGGTAGCGCA GCGCGCACCC AGAGGCTCCT GGGACGGACA CCACCCACAG
CTCACGGGTC CTCTGGCGGT GTCCTTTGAG CTGCCGGGAC GGCGCAAGGT
GGCTGAGCGA AGCTTTTTGC TCGgtgctta cagggagtac ctgagatggt
ttgggaaagg aaacaggggg tctcagaccg cttgggtgca gcgctctccc
catgcgtttg agtgcagatc tgtcgagacg cccagccct cgcacccca
ccaagctctc aggtagagag gccccgggac agggtagacc ccatgtcct
ttggtgggca taacacagag ggaggagcaa ttgtttacac atgtgccag
tacaagaagg gaaactgata tttaagactg aagtggggaa acgggagggg
aaggagccct tatcctggat gttagacgct attaaggacc atgctatttt
gtatgtgtga gaatgatatt gaaggtaagg aggtccttat ccttgaagat
gcatgctgaa gtgtacagaa gtgaaatgtc acgatgtctg cagttttcaa
atgatgcagc taacatacgc acacattatg agagagagaa cagaaaggaa
gcaagtattg caatatgtta acagttgttg catcctgttg ggggtggggg
gtaaacagac gttcgttgta ccgttctttc aacttgccca aatatttgca
aaatgccata atgaaaatth tggagaggtt ccctataaaa atagtgcctg
aatgctaaga aatgaacgc ggcttctttt cctttccctg tcccggtggtg
gacaaatctg cgagcaagca gggctcaagg ccagcggggt ctggacggga
agagcaatgg gtagagccgg cctgcagctg ccatagccct gggcaggcca
ggagggggcca tcggagcctc ggggcagata ggctgcccgg cgcaaaacta
tgcgaggagg cagccgagcg cgttggcccc gggggcgggt cggggtgggg
acttcttggc ggctctcagc aacccttggt ccgaccagac tcggtgggct
cggctgcacc acggatgagc aaagaaaccg gcgggcccgc ggaactgagc
cgcgcgcctg cagccgtcgc cacggcgggc cgcggccttc cccgcattc
cccgaggagg gccgtagttg tcgcggcctg cagacagcgc gcagggtagc
ccagaaccgc gtgcggggcg tgggggaaaa aaaattaggt gccgaaagaa
agcagcaacg cacgcgcgcg cacacacaca cgaaatcagt ttacaaaaat
aaaaccgcgg aatctgccgt gctcggcgcg cgagtttgag ctacttaact
ctcgcagca ttttactgtc atatttctgt agagctgatg gaaaatctga
aatgtttttg caagggcgat ctgcagacc caaggcgttg caagcgttgc
gcagcctttc gctggcgtct gcggctcttt tccggggcgc gaggagggaa
aagagtagtg tctttcaggg tctgcaaagc tgttttgaaa agttaaccac
aaaggtttta aatagcaaga ttttcttttc aggaaaagtt tttaaaacgg
ctgatactcg tccagttttg gtgggcgcta aaaagtctta ttactaagct
acaagtttcc ctgttaagat tttcaaaacc cctcgtggac agggctgaaa
gtttgccggg ctgtgagacg tgactgaaca cgcttctaaa tctcagata
gatcatttat aaagctctgg tgcccttggg atgaaaggct tcgctgagtt
tttaagactt gccttcgtgg gttaaatttgc ttgggttgct acacgtaagc
attcatttcc ccacctaaat tagtaaaagg gcagaagatt ccaacgacgt
aagataaaag ttgggtttat ggcagtttct ctaaagacaa attgcaacag
ctttctttat taaactcctt ttgctacctc agcaaattgg agagcagtt
gaattattgat gaattctttg agggtagta atattgctac aggcgggaga
aagtaaggga ggacaagctt aagcatttac tcaagtgaac aaccttcaaa
tgtttttcaa ctatgatgtc tcaaggattt tcttaggacc tttgaagcaa
accctgattc tgcaaatggc ttcttatttg gtaggcctt cccaacctga

```

Continuação:

caaaactcac	aggatataaa	ttgccctaag	aagtgggtgag	aaacttaccc
tagcattgga	aacaggcaca	gcccaccacc	tgggggtgta	ggggggcctc
caactggtag	tcaggcaact	tctcacgggg	caactaggaa	cctgaaatga
aaattgcaca	ccatctattg	tttttttttg	tcccttctga	aatattgttt
caaggcagct	gttttcagga	aaatagcaca	tatttttcctt	tctctttttg
ttccgtgaaa	cgttcggaaa	ttatattgat	aaggaataat	catagtccgc
cgagggagca	atagctttga	aggctctgaa	ggcacagctg	agaaaaataa
attagctcta	gactagagat	actcagcaag	gaagttatgc	ataacatggt
atlttggcata	aatttttgaa	aatgttttct	ggaaagcaac	tgtgaatgaa
tcttgggaaa	tagatactgg	agaatcaagt	tgcattaaaa	cttatttttt
tgtaaattac	aaagaagatt	tctagaactt	tccattttta	ttttaactag
aaatgggtgtt	taagattgtg	tttttaaatga	aggccttgag	ttagggaaac
aagattttcc	ttatgttgta	aatatccaga	agtctttaga	gagataaaaat
caggatagag	ttacaaatth	aacagaaaaga	aaacaactgg	aattgtgtta
tgaagcttct	attgtgtttt	tttgtaaagg	ctttttttca	tggttcaagc
agtggctagc	agtgagtagt	aactgtatac	tataatataat	aatgtaagaa
tgcttgtcct	actctcagat	taaacaaaaa	agacagtgtc	ctttggtaaat
ttttattttta	tgaaatttagt	tataaaaaata	gttattttgca	aaactcagaa
ctttggggaga	aaaactctaa	tcttaacttc	tcctaccccc	gtgatcccat
caggaaaaag	tagaatgaat	tagtaaatth	ttcctttctg	ccaaagagaa
gcttagtaag	tcattagtag	ctgttgcttt	tgatttctag	atctcttttc
cctcctcaaa	gacgcagaat	cactgccttc	agtgtctgct	ggatttcttc
tttaaaccag	gctttttttt	ttttttttct	gggtgtgtac	tattattgtc
ttaattattc	ctggattttt	gcccattaaa	tcaaaaattg	tctcagatga
cggcaatgat	tggaaacttt	tgctaaggag	ggcgagaatg	aaggggtgta
ggacagtctc	agcgtctctt	tacaccgtgc	tggccctata	gtgcaacctg
taagtgggca	ggacttgctg	tcctttgcac	agagcgtgaa	ctgacagagc
aaaatgtaat	gacaggagca	tggagaccaa	aaacccatct	cttgtagggtg
tggtggccat	aatggcggtt	caagtatagt	actaggctag	atthtgcac
ttgctttcac	cataaaaatta	tggtaatcag	tttcacctca	cacttctaga
actttcctcg	aacttcagct	aattacgaaa	aagagaaaag	aaaaacccca
atataatcagt	gtcaacctgt	caaatacgaat	cttggtgagg	gaaactgaag
tgcttttgga	aggcacttca	ttgatataca	aaaaattgaa	acaaacaaac
aaaaacacca	acaaaaagtg	atcagcacca	aggatatcct	gtaaaaagca
cagagctctg	ctttgtgatg	tctgaatgac	atcgcttttg	ggatcttggt
atgtgaaaag	aaaacactcc	tcggggcaaa	taatatcttg	ctgggtagaa
agaataggag	ggaacaacaa	attgaagagt	gctaagggat	gctatccttg
ttcaaaacca	agatgcagct	ggagtgtttt	ccaaagcatg	aaatacattg
ttgaaatagc	agagagaaaag	tgaatgtgca	tatctgaaat	gttcacatga
gagatggtaa	cacggcatca	tccagaactg	aactcactcg	tggttgcoat
tctttccctt	gaagctccgt	ttgcattttta	aaaaataaaa	cttaagagat
agagtgtcaa	ttcgccgggc	atgggtggtg	atgactgtaa	tcccagctac
tcgggaggct	gaggcaggag	aattgcttga	acccgggagg	tggagggttg
agtgaaccga	gatcgacca	ctgcactcca	gcttggacac	gactctgtca
aacaaacaaa	caaacaaaaa	aaacagcgat	agagtgtcta	ggactgatgg
ataatgtcat	ttcttttgct	tcctgtttct	taaaattatt	cttgggctta
cgatcggtg	ggattttgca	gttcatgacc	agagtaagtc	ccagccctcc
gtgtagaaga	ggaagtagaa	atccagatcc	tgtttttaat	tttagcattt
ctcctaaacc	cagggttga	cagcttcttt	agtccagagt	ccccaggaat
gcctacgcca	taggcaggga	agagtgaata	tgagagtagg	gtggagagca
agggccaggg	cagaagacag	aaagaaagaa	gatgaaagaa	gagaaaaaaa
aaaacaagtc	cttctcctct	gtccctacta	tctgtcctca	agttcagagt
ctccacttgc	caccatgatg	tcatcaaggt	gaccttctca	ggccccccac
tgatctctca	gttgagttct	gtcctctctc	cagtgggtcca	cggagtctctg
ctgggaaaac	acaattcttg	tcatattaat	cctcgggcaca	aaaggcagat
gactaaaaag	cttaactcct	taatgtggca	ttcaggggcat	tctacaatct
ggctccaatc	tgcgcttctg	cttgcatctc	ctaccctcca	gccacaacct

Continuação ...

ctttatgtgt	ttatgctgga	tacccctggt	cctcccattt	ccgcatacctt
cttaggccca	gctcaaagtc	cacctcttcc	ttgaagcctt	ctcagatggc
ctcagagttg	gagttaactg	cctctgtcct	gtcctgtcac	ggcactttta
ttaacttagt	tgttctgaaa	ttctcttcct	agcagacagt	aagctcctag
agggtagggg	catgtctgtc	tataggtatc	ttctcagcgc	tttgcacaat
aatttgcac	cgttgatgtc	caatacatat	ttgttttttt	gttttgtttt
ttttgtttgt	tttttgagat	agaatctcac	tctgtcgccc	agcctggagt
acagtgggtg	catctcggct	cactacaaac	tccgcctccc	aggttcaagc
aattatcatg	cctcccaa	agctgtgatt	acaggcacgc	accatcacac
tcggctaatt	tttgtatttt	tagtagaaac	ggggtttcac	catgttgacc
aggctgggtc	caaactcctg	acctcaagtg	attcacctgc	cttggcctcc
caaagtgtcg	gaattacagg	catgagccac	tgtgcccagc	caagttagaa
agaaatccaa	tacataattg	ttaaactgaac	gtatacctga	gaagggaaaa
aaggaaccgg	aagaagtgag	aacaaggtgc	tgtgctggcg	gtaaagatga
tggctaacat	cctcaagagt	tggcttaggc	accatttggg	gttcttgggt
ctgaccaagg	gttgttgagc	tcatacttaa	catgttggct	tcacaaacat
ccatctggga	ctgacacttg	gaacacatcc	tctggaataa	tatagaatta
atltgtcttt	aaattagctg	ctcagtacta	taaactctct	cctgaaaatc
tatcaccatc	atltttatact	tttaagcttc	aagctgatcc	ttgaagggaa
gggtggaatt	ggatggaatg	ggcaaaacag	catttttcaga	ggccaggagg
cagagaagaa	ctgaaaggag	atggacctag	ttggacagag	cctttctact
ggggaattag	agactgggct	ggacaagtag	ggtgatgtca	gattattggg
gggtggggag	gcaggcgcg	tggctcacgc	ctgtaatccc	agcactttgg
gaggccgaga	cgggtggatt	gcttgaggcc	aagagttcaa	gaccagcctg
gccaacatgg	tgaaaaccca	tctctactaa	aaaatacaaa	aactagccgg
gcatggtggc	gtgtgcctgt	aatcccagct	actcgggagg	ctgaggcagg
agaatggaga	gtagcttaaa	cccagggggc	agagggttga	gtgagtggag
atcgtgccac	tgcactccag	cctgggtgac	tgagactgtc	tcaaaaaaaa
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaagaggctg
ggcatggtga	ctcacgcctg	taatcccagc	actttgggag	gccaaaggcag
gtggatcacg	aggtcaggag	ttcaagacca	tcctgaccaa	gatgatgaaa
ccccgtctgt	actaaaaata	caaaaattag	ccgagtatgg	tggcaggcgc
ctgtaatccc	agctactcag	gaggctgaga	caggagaact	gtttgaaacc
gggtagcaga	ggttgcagtg	agccgagatt	gcatcactgc	gctctagcct
gggcgacaga	ctgagactct	gtctcaaaaa	aaaaacaaaa	aacattatgg
ggggtctgga	agtaggcaga	ggagttgtga	ctttagttgc	aggatgggat
agaacctttt	tttttttttt	tttttttttt	gtgtgatgct	tttaaagggg
caggggctgt	atgtaatgaa	agccgtgttt	tttaacagat	tcacctggct
tatatggtat	aaccctatga	actacaaact	atggtaaacc	taggactatt
tgaataatta	gaatatccca	ttctccatgc	cacactggag	tgaactttct
tatttgagaa	gataaagagt	ggaaaaatac	tatttatcat	atacctatta
tgtatcagga	gatgtactgg	attcttcccc	caccccagg	attccctccc
tctctcccc	ttccctccct	ccctctctca	cacacttgca	cttgacacac
ccagaaatcc	caactaggac	ttgattttta	ttgaccttta	aagtaataca
ttttttgtta	aattactcat	acaaagcatg	ctatgaaata	ccagtaataa
tcgggttaaa	aattaaactga	aagacagaaa	cttactattg	tgtgctgatt
tgatggaaca	gaggagtga	aatatagtac	gcaccttctg	gattgagact
gtggttttcg	tttttgtttt	tgttttttgg	tagtggcaga	gcttcaaact
ggagaatgtt	ctttttcatg	cgcagttttt	aaaaattgtc	attaaatatt
ttcacaaggt	taaaacaaca	ttgaaagtaa	cttgggcaac	ttagatatcc
cctgtacttg	ctgtaactca	aataccaaaa	catattttaga	gttttaggtgc
atctatcatt	cgaggcatgt	gatttcttta	cctgggtgaa	gcgaaagaaa
gttaaactga	tgaatatatga	aatgaaaatg	agaatcaatt	tgtagaatga
accataaaag	aacgagattt	ttgaagagat	tattttattt	tgattgcaga
ctgtaaatga	tcacagggag	ttgctgtagg	aatgacattg	aggcttactt
ttgggtcaat	ttgttcctgt	cctgaccac	ccttggcagt	ggagacagct
agaattttct	taactccata	gaaaccagtc	tcactgttgc	tttaggactc

Continuação ...

```

gattctctcc ctacactggg cttccttttag tgctttgtta tgtaaataca
gcttcaaaaa tagaataaat attctgtagg cattacttgg aaaacgagct
tttattttta agatgtgtgc ttatttaagt cgcagacaga gctttaaaga
gctcttgata gaaagaaata tatttggcga agagtcaaaa cttttaaagt
aactttaaaa ggtgaataat ttacattcgt atagtgttta cataataggg
tgtgtacgac aaagccaaaa caacattgaa aataacctga gcaatttaaa
tattcttcct agtaactgta aaataaaaagt gccttcgaat ggtacatttg
ctggcaggaa attttttttca tgttcccagg tggcaaggat tctttcatgt
tccccttatat ttgaaaactg gtattagcgt ttacttctcg gactcagacc
aattccccct tacctgtcag agaacatcta gcaataaatg ctgacaaagc
tacatagcaa agggcaaacc tgtcacttcc ctttcagaaa cctgtttatg
caccctcctg caatcccacc acctgcccag gccaaagtga cccctctccg
ctactggtgt ttcataatgc tgggagctac tttgatttca tgcattgcc
ttcaggtaag ccacaggaac ctcaccactg gccatgccag atgggtgtga
attggataag aacaaagaaa aatggaagta gcaagtgggt gggagttag
tgtactatat gacttactca tatgcaagtc aataagcaaa gtcccctgcc
tcttgctcct ggatgctgga taataataaa tataaatata cataatatat
atgcgtacat atatacatgc atgcacacac acacagacac acattacctg
gatccccac ctctcacc cactttgaaa aacttttagag tatcgttggc
ctcatcagaa ttacattaca ttagaggatt gttgaaagat ttcattattac
tgctaattgt gggaaacataa ttagaaccaa cgttagcgcc tctactcagt
atcatcattg ctattatattt atttacatct tatgcttttt tcctttggct
ttaactgtct tcatcaatct cctgttataa aaataaataa aagcttcccta
tggaattatt ttcaccagtc aagtcctgta acttgttttc cttgaagagg
gaaaatataa ctgatggaca caaggaaaca gcagtggcat agcatgtcct
tttccacca gtctgtgttt tgacatcagt tgtatttaca cacataaact
gaagaacggt gcatttgtgg agtttgccag tgtggttaaga ttgaaggggt
ggtaggcagc ctctatgctg gtccccactg atccctgcct tctggtgttc
ccactgttct gtaatccctc ctagtacgtg cgctggacct agagactcct
ctaacaagta gaatatggca gaagccatgg gatgtcactt ctgtggcttc
taacttgggc cttttgtcat ccccttcccc ttctctctat cagCTGCCAT
GTTATGAGGA CACTTGGAGA AAGCTACAGG AGAGGAACTG AGGTCTCCTG
CCATCAGCCA CTTGAGTGAG CATGGGAGTG GATTCTCCAA ACTCCAGTTA
AGACTTGAGA TGTCCACAGT GGCAGCCACT AGCTTGA CTG CAACTTACGA
GAGACCCTGA GCCAGAGGCA CTCAGTTAAG TCATACACAG TTTCTGACC
CGTAAAACCT GTGACACAAT GAAGGTTTGT TGTTTTGGGG GTAATCCGTT
ATTGAGCAAT AGATAACGAA TACAGAAGGC TTGTAATTGT ATAACCAACG
TGAGTTTATA AGCGGATATC TGACCTCATT TGTTTTCTCC TGAAAAAGTT
ATAGAAAAAT CACAAGACTG CAAGTCACTC TCTCTCTTTC TTTTGTGGGG
AATAAACAAAG TATAATTACA A

```